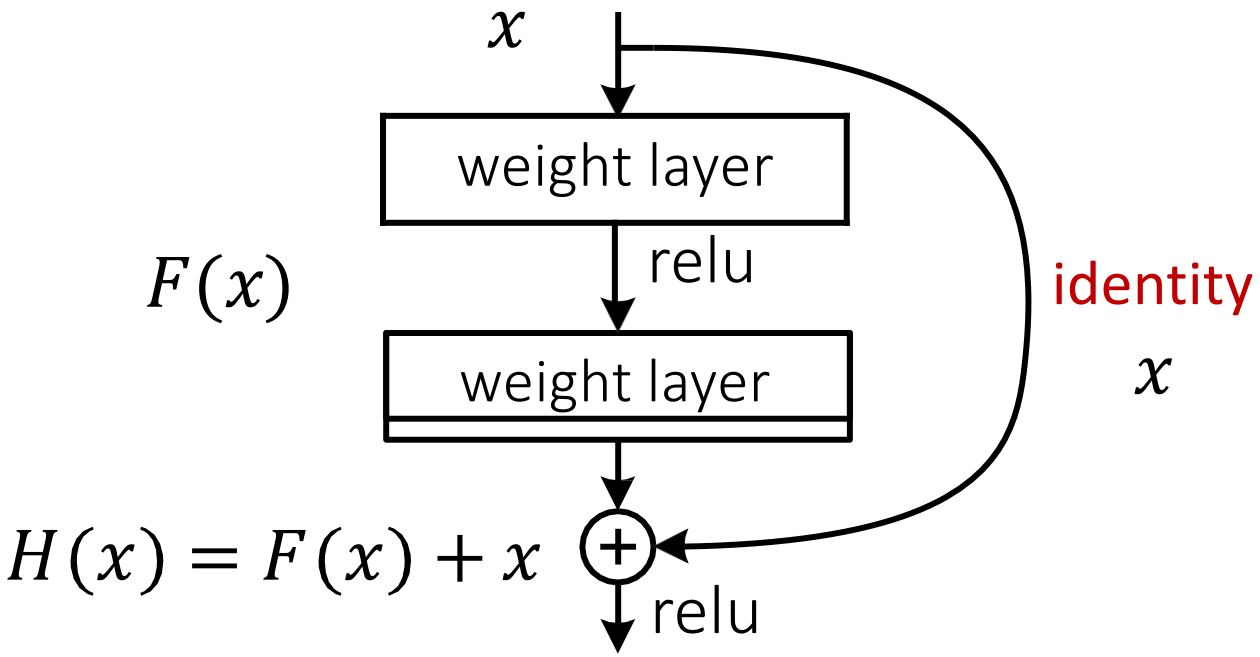
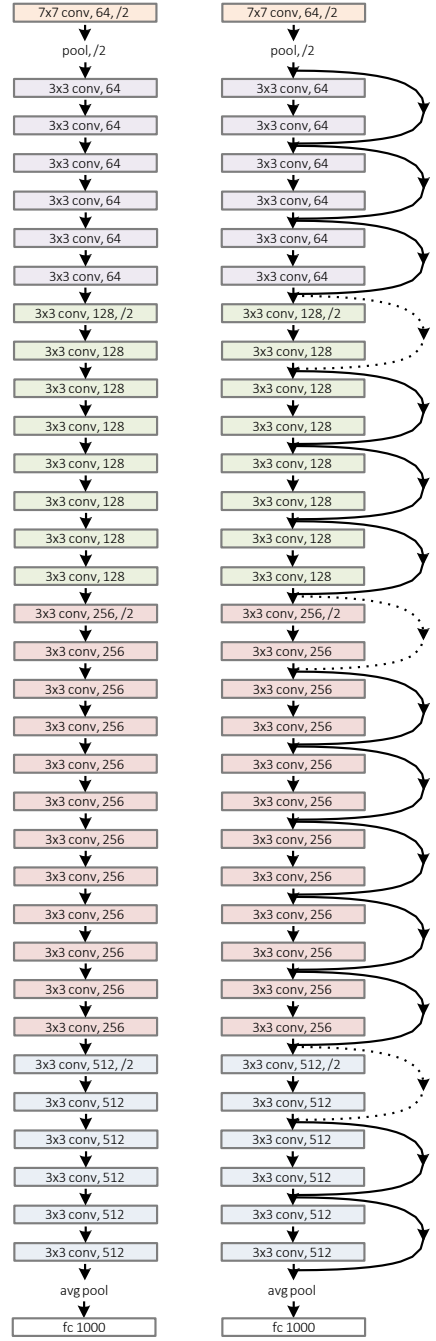
Recap: Resnet

- $F(x)$ is a residual mapping w.r.t. identity

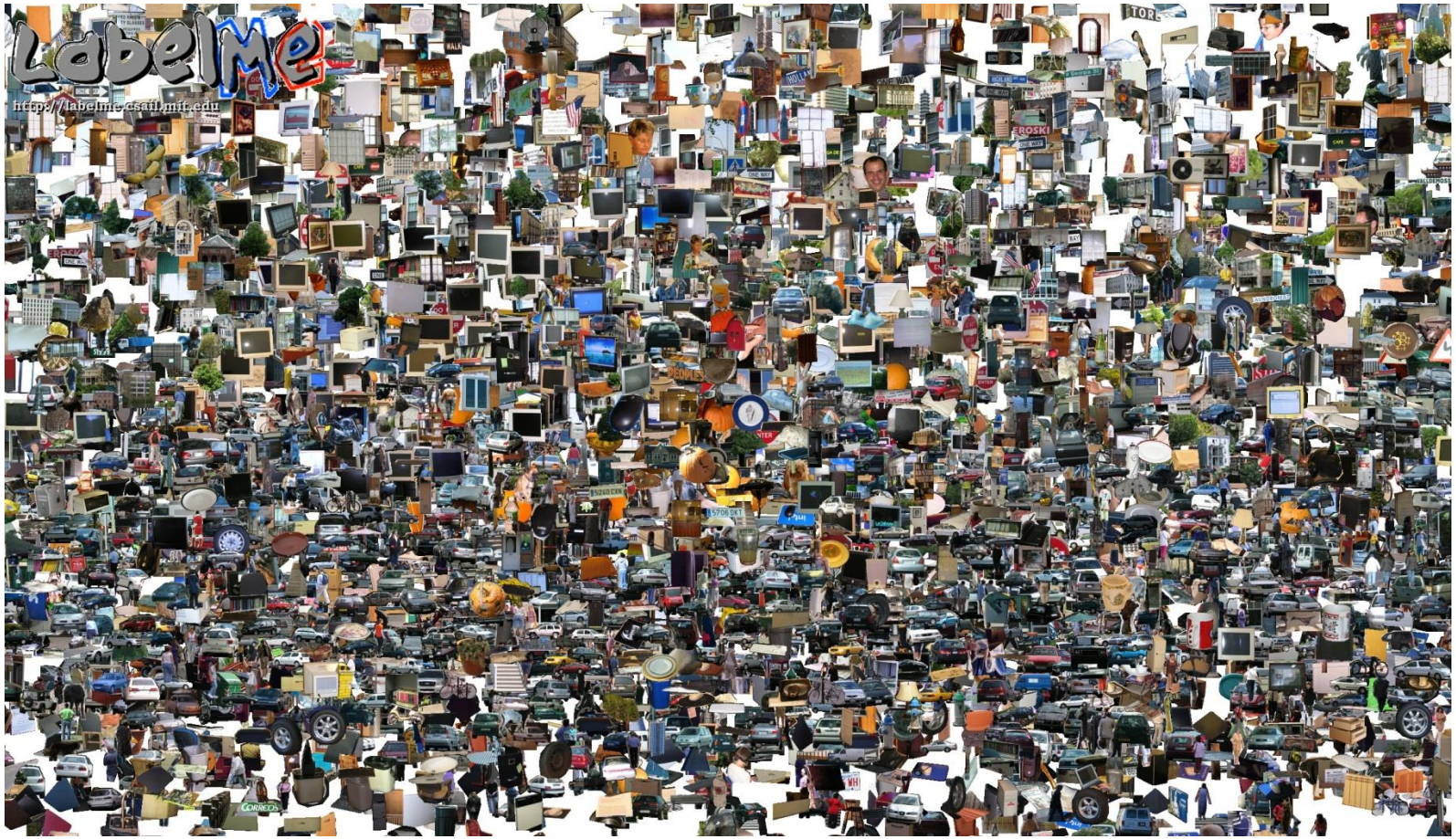


plain net



ResNet

Opportunities of Scale



Computer Vision

James Hays

Many slides from James Hays, Alyosha Efros, and Derek Hoiem

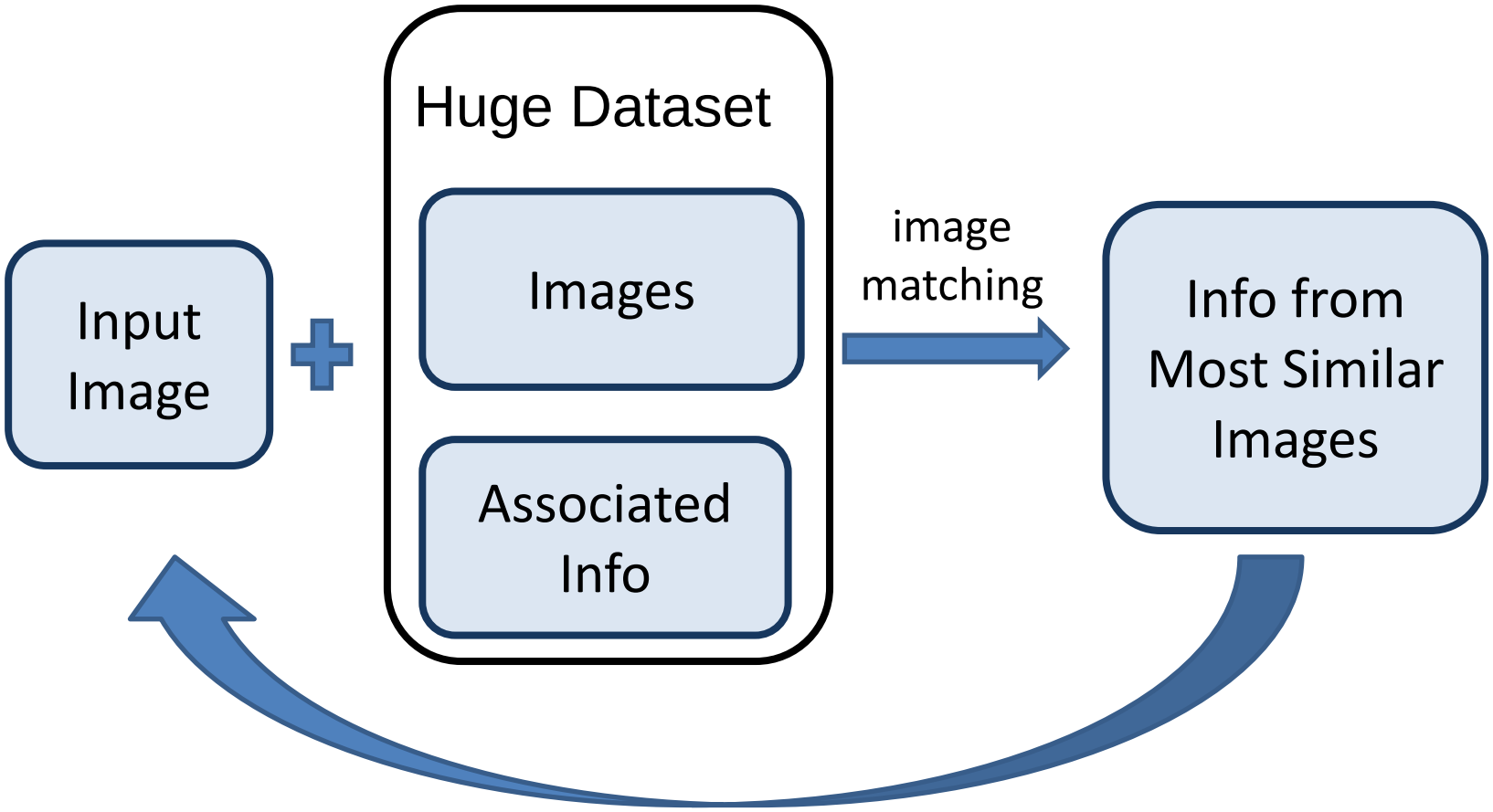
Graphic from Antonio Torralba

Outline

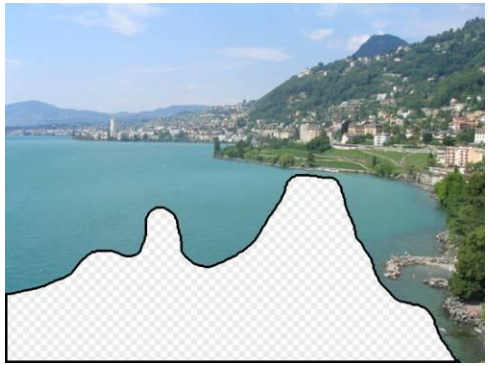
Opportunities of Scale: Data-driven methods

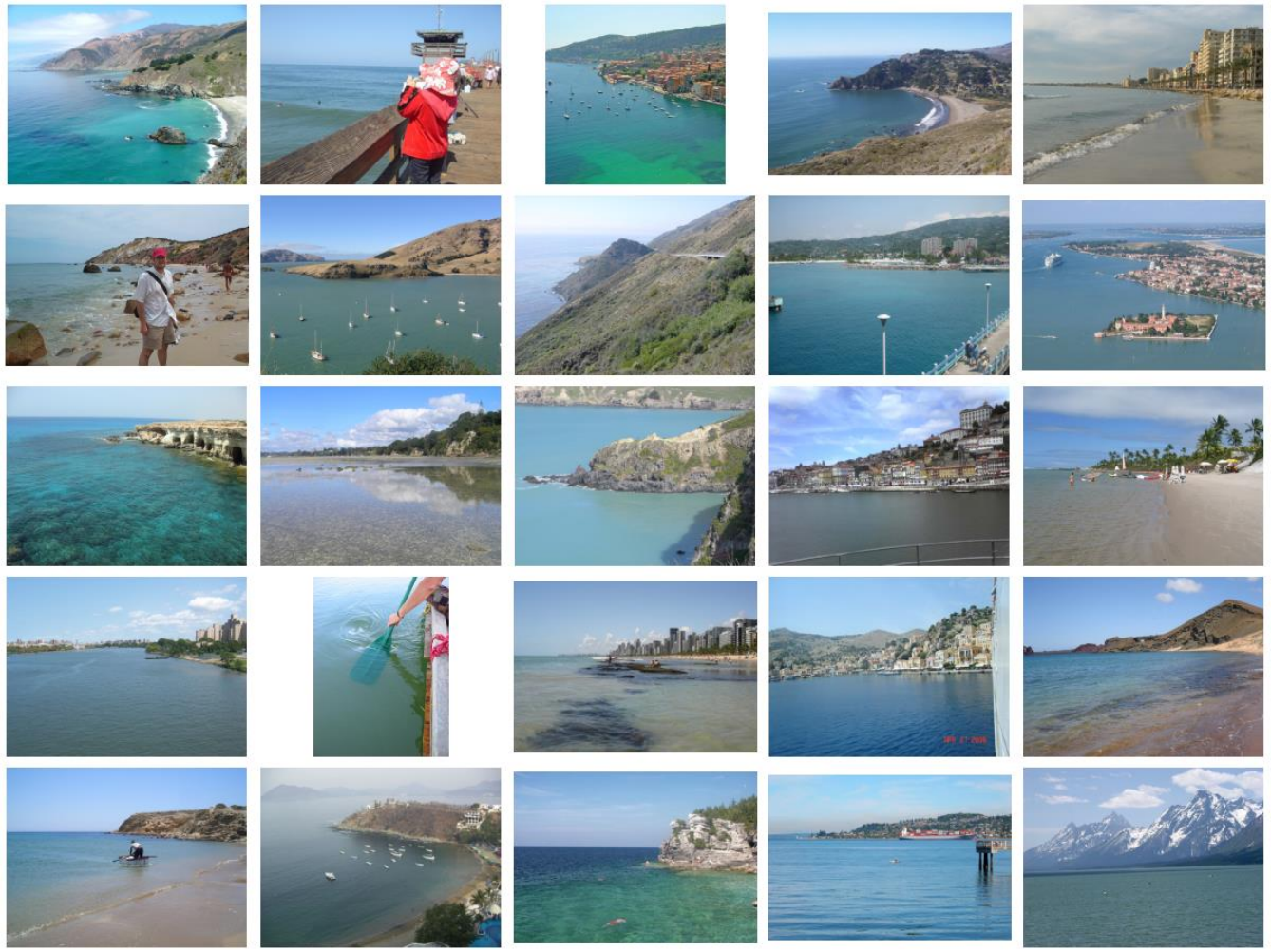
- The Unreasonable Effectiveness of Data
- Scene Completion
- Im2gps
- Recognition via Tiny Images

General Principal



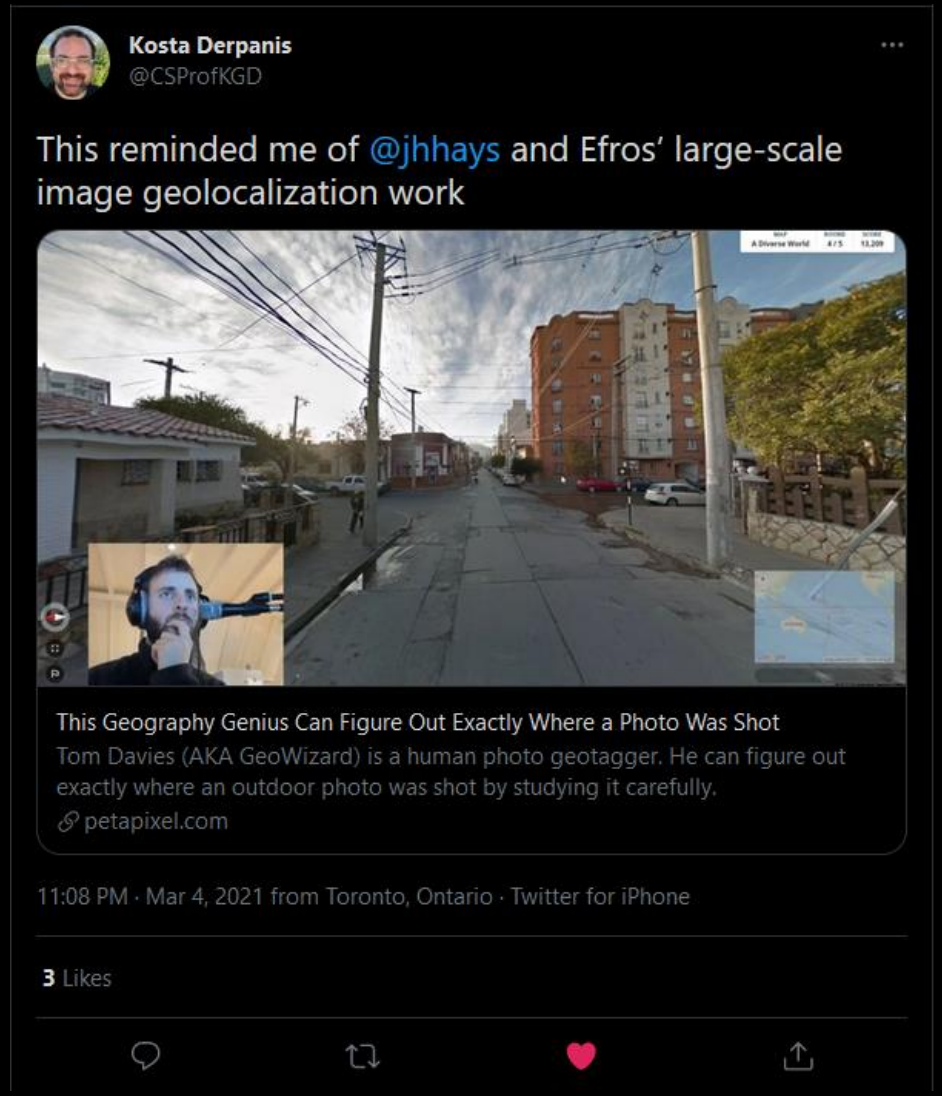
Hopefully, If you have enough images, the dataset will contain very similar images that you can find with simple matching methods.





… 200 total

Graph cut + Poisson blending



https://www.geoguessr.com/

https://www.youtube.com/c/GeoWizard/videos

im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images

http://graphics.cs.cmu.edu/projects/im2gps/

How much can an image tell about its geographic location?

Paris

Paris

Paris

Paris

Paris

Paris

Paris

Madrid

Rome

Paris

Cuba

Paris

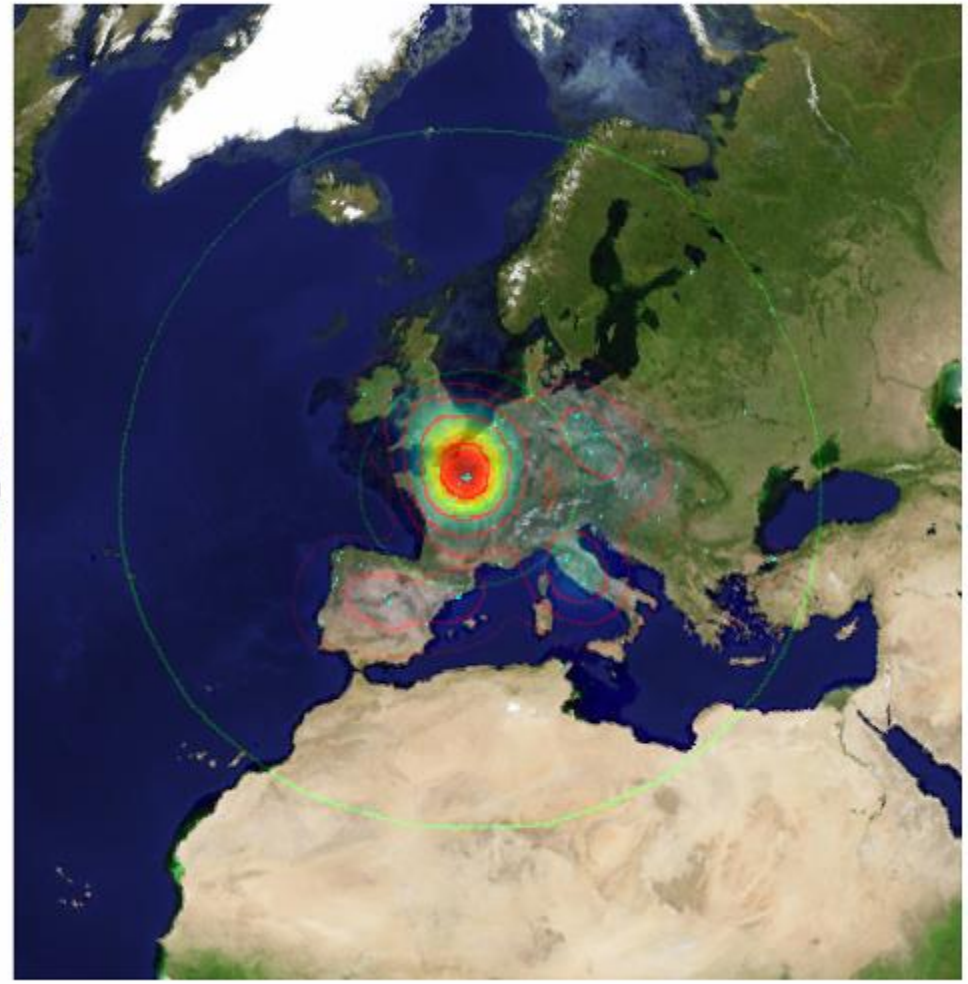
Paris

Poland

Paris

Paris

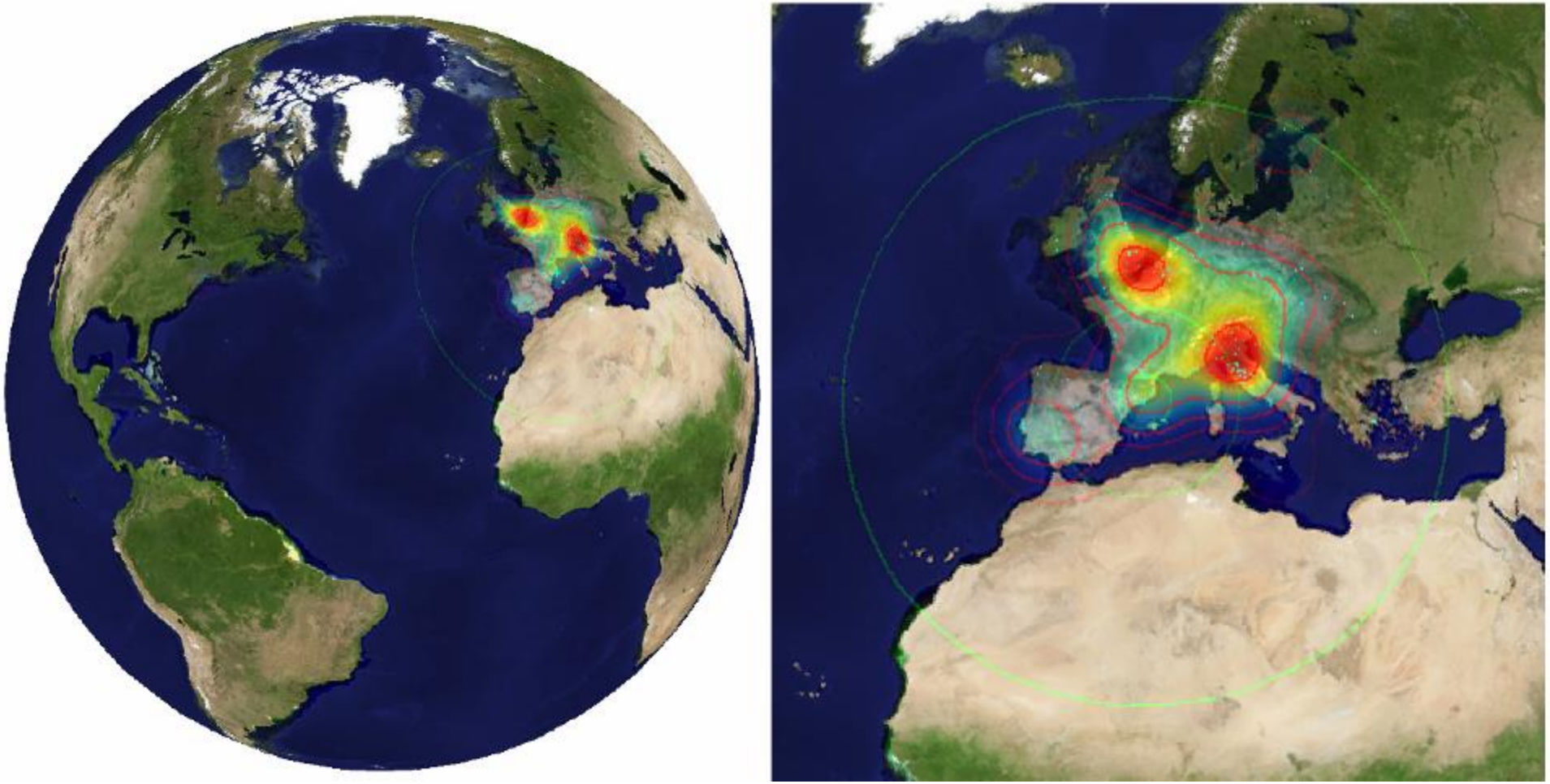
Nearest Neighbors according to gist + bag of SIFT + color histogram + a few others



Im2gps

Example Scene Matches

Voting Scheme



im2gps

Philippines

Houston

Thailand

Houston

Maldives

Philippines

NewZealand

Bermuda

Palau

Mexico2

Brazil

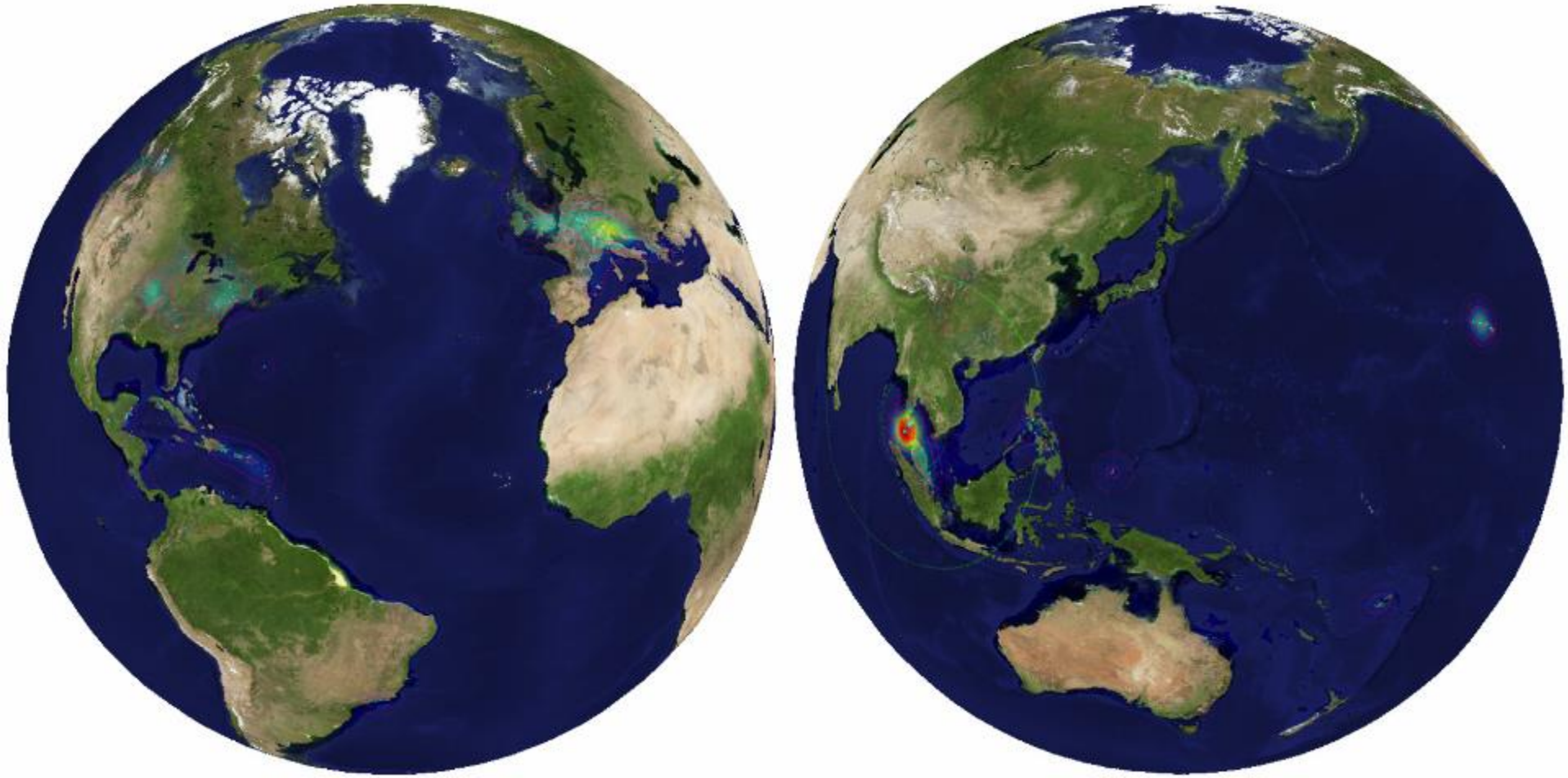
Mendoza

Brazil

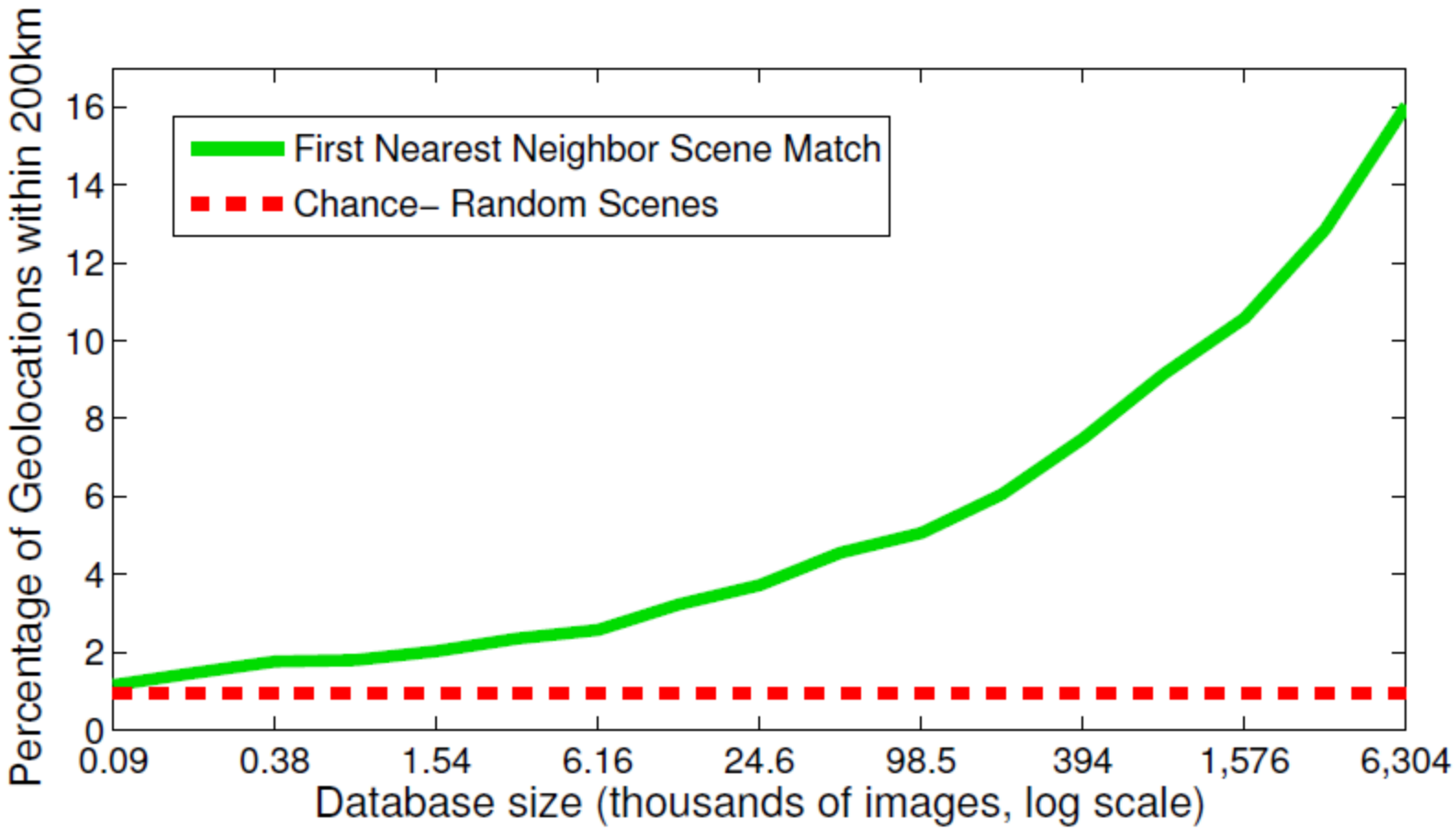
Thailand

Arkansas

Hawaii

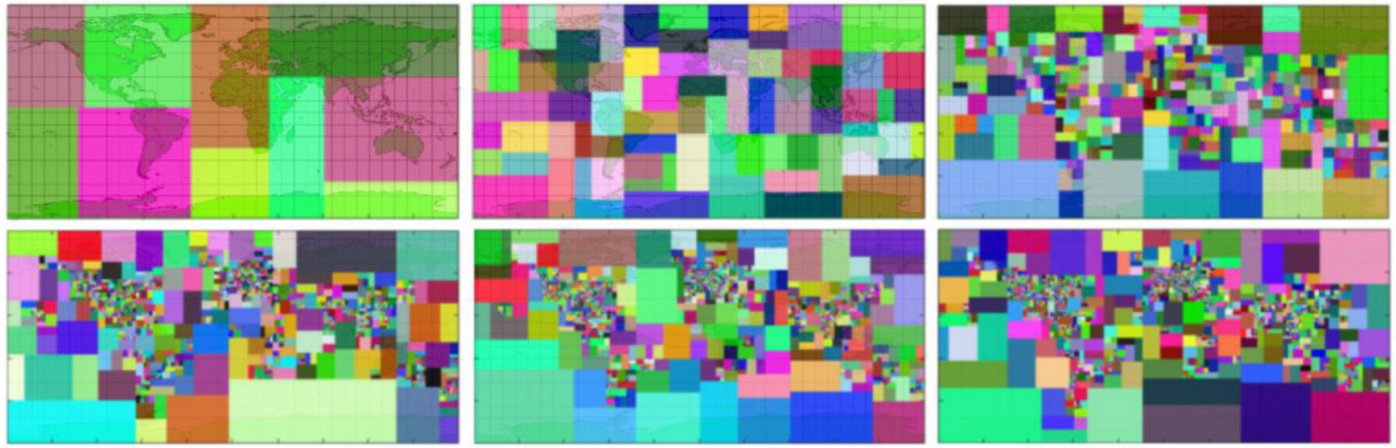


Effect of Dataset Size



Follow up works

- PlaNet - photo geolocation with convolutional neural networks. T. Weyand, I. Kostrikov, and J. Philbin. ECCV 2016
- Revisiting IM2GPS in the Deep Learning Era. Nam Vo, Nathan Jacobs, James Hays. ICCV 2017



Threshold (km)	Street 1	City 25	Region 200	Country 750	Cont. 2500
Human*			3.8	13.9	39.3
Im2GPS [9]		12.0	15.0	23.0	47.0
Im2GPS [10]	02.5	21.9	32.1	35.4	51.9
PlaNet [36]	08.4	24.5	37.6	53.6	**71.3**
[L] 7011C	06.8	21.9	34.6	49.4	63.7
[L] kNN, σ=4	**12.2**	**33.3**	**44.3**	**57.4**	**71.3**
... 28m database	**14.4**	**33.3**	**47.7**	**61.6**	**73.4**

Tiny Images

80 million tiny images: a large dataset for non-parametric object and scene recognition Antonio Torralba, Rob Fergus and William T. Freeman. PAMI 2008.

http://groups.csail.mit.edu/vision/TinyImages/

256x256

256x256

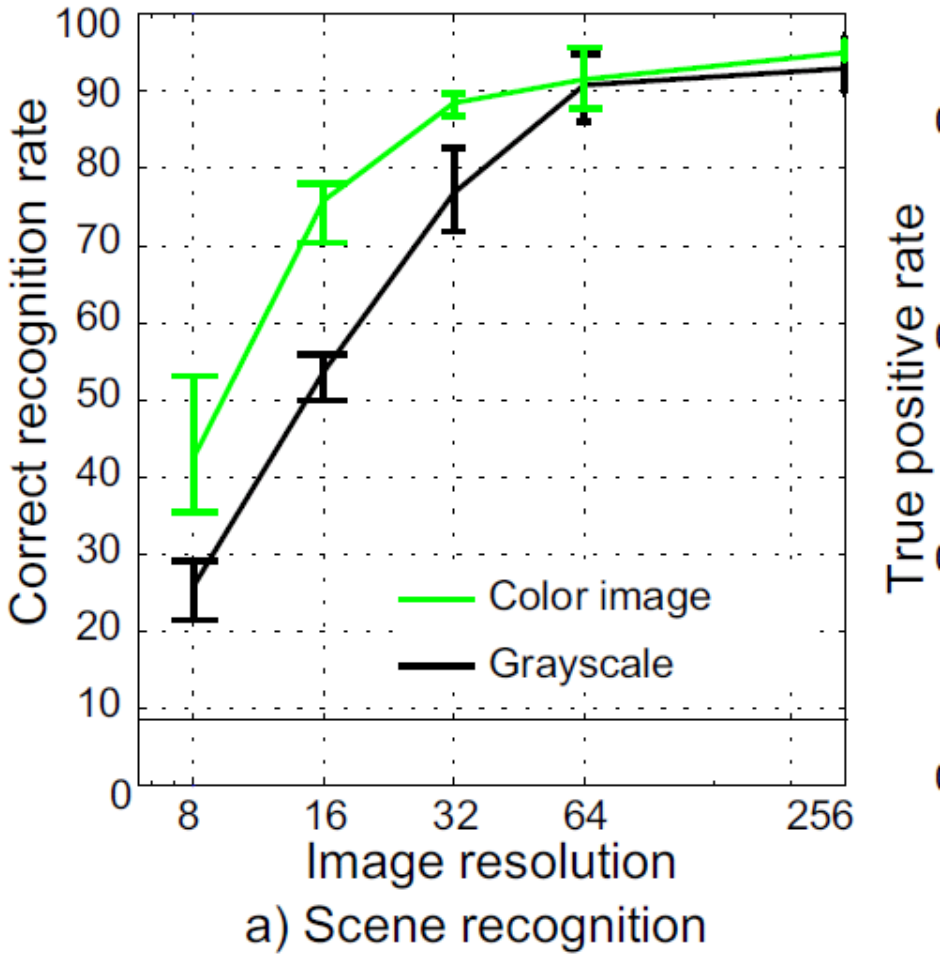
32x32

office | waiting area | dining room | dining room

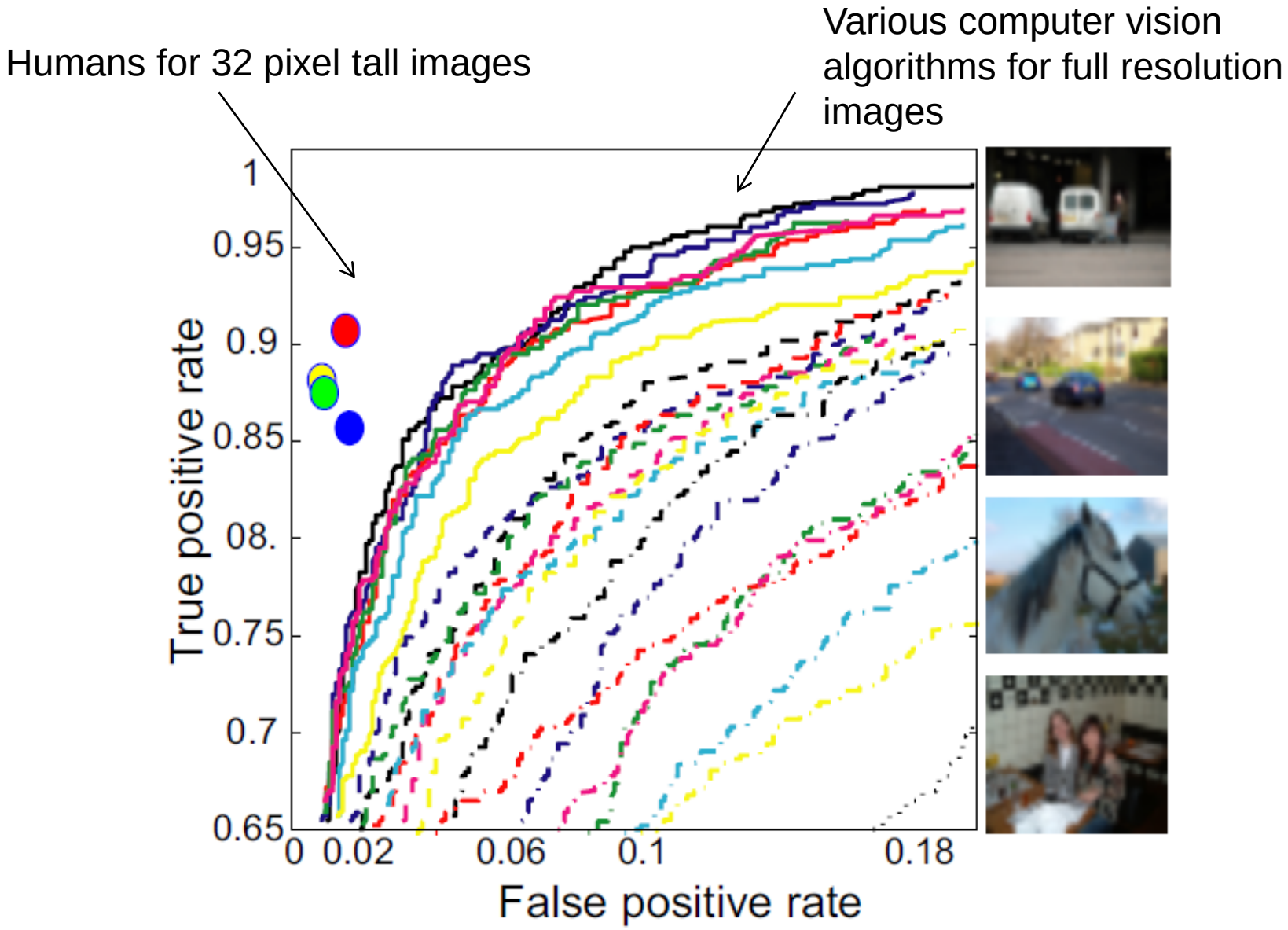
c) Segmentation of 32x32 images

c) Segmentation of 32x32 images

Human Scene Recognition



Humans vs. Computers: Car-Image Classification



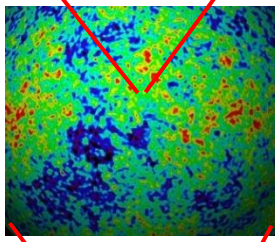
Powers of 10

Number of images on my hard drive: 10^4

Number of images seen during my first 10 years: 10^8
(3 images/second * 60 * 60 * 16 * 365 * 10 = 630720000)

Number of images seen by all humanity: 10^{20}
106,456,367,669 humans[1] * 60 years * 3 images/second * 60 * 60 * 16 * 365 =
1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of photons in the universe: 10^{88}



Number of all 32x32 images: 10^{7373}
$256^{32*32*3} \sim 10^{7373}$

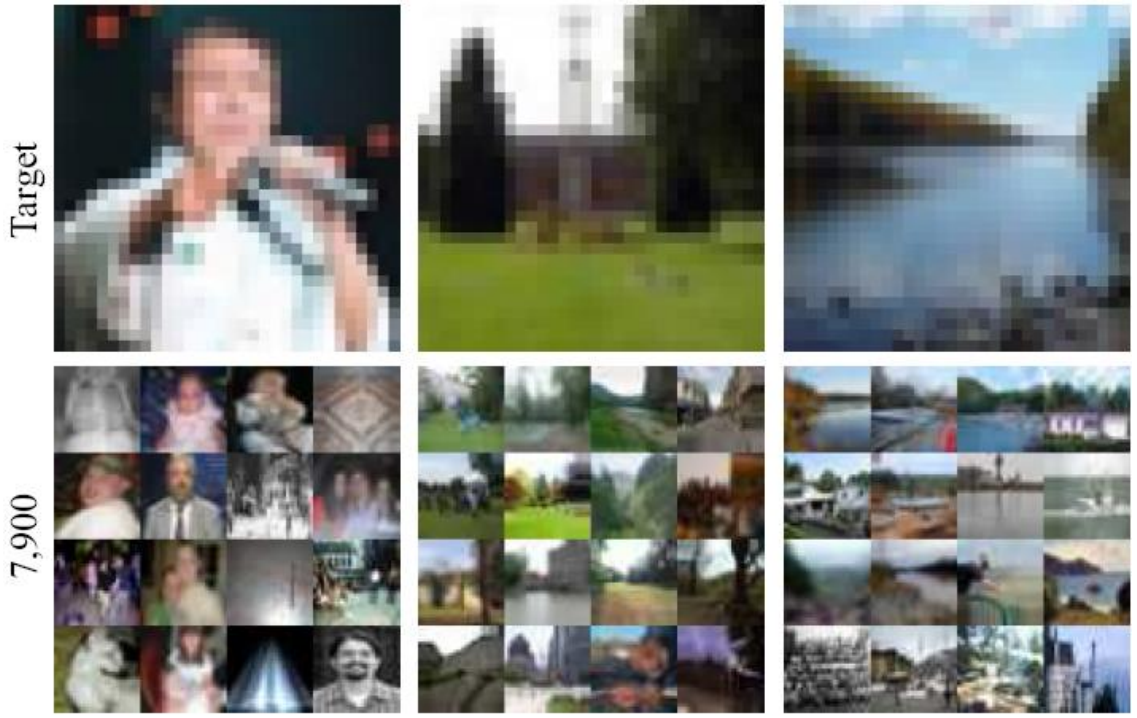
Scenes are unique

But not all scenes are so original

Lots

Of

Images

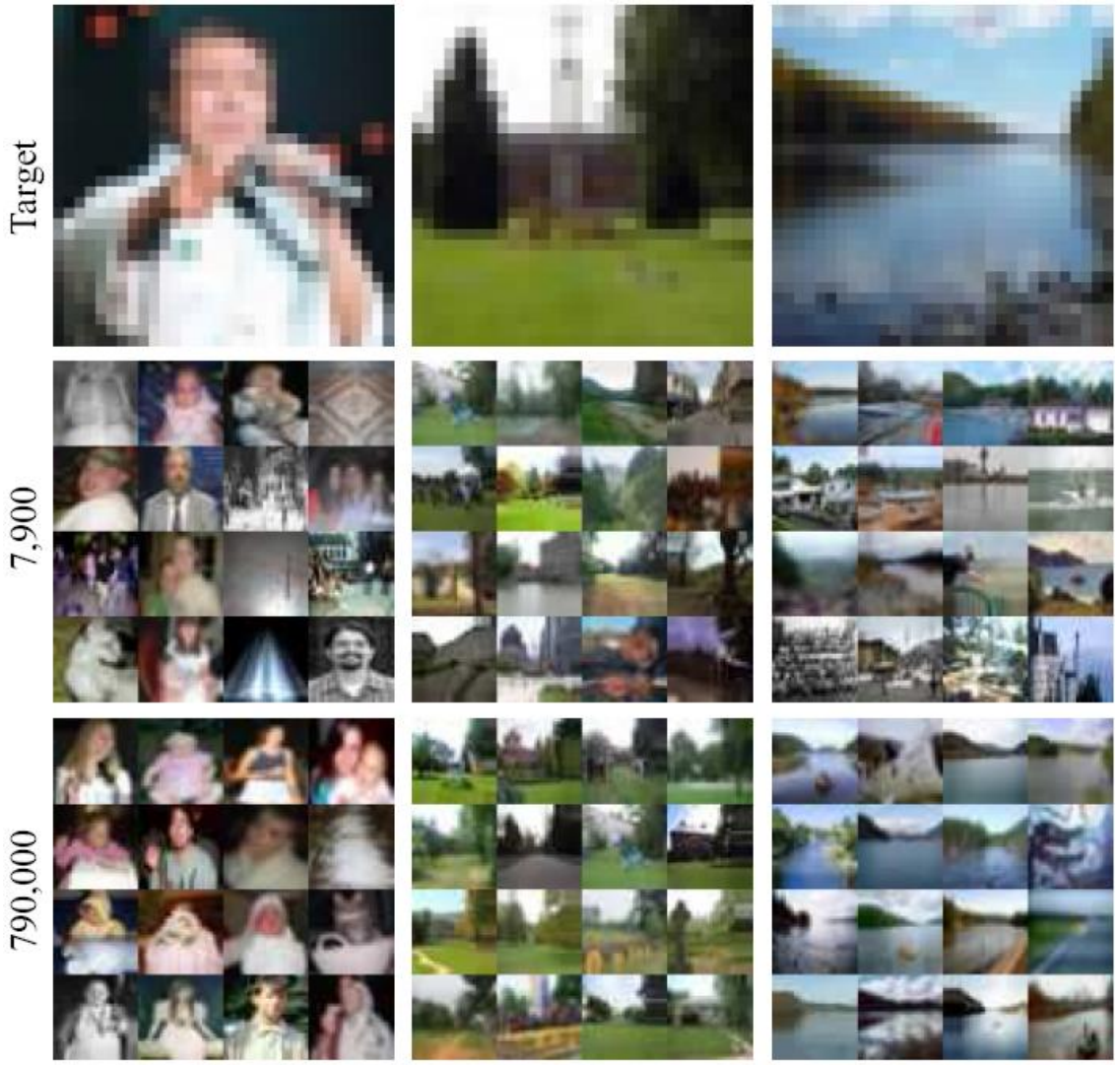


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images

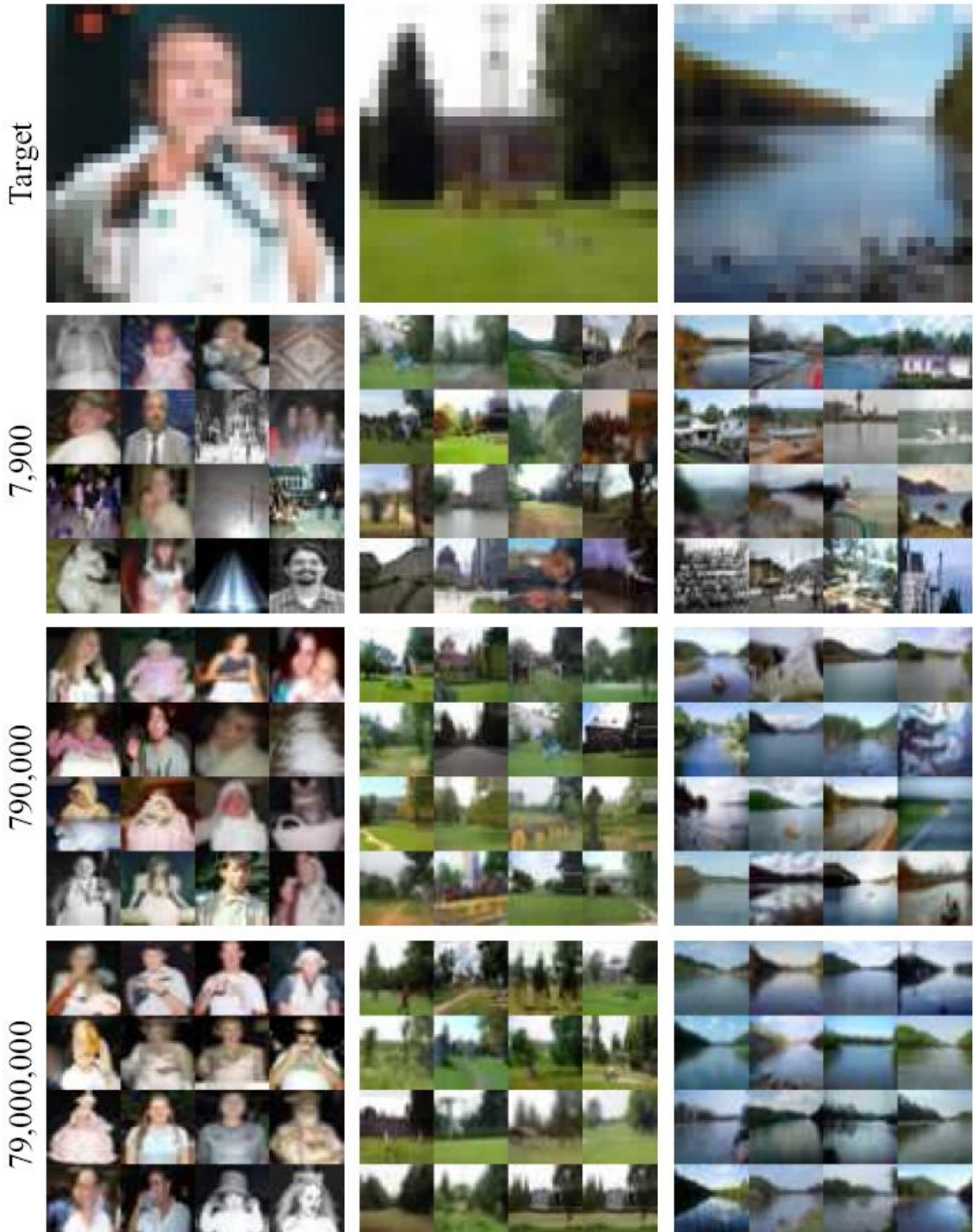


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images



Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match

Application: Automatic Colorization

Input

Color Transfer

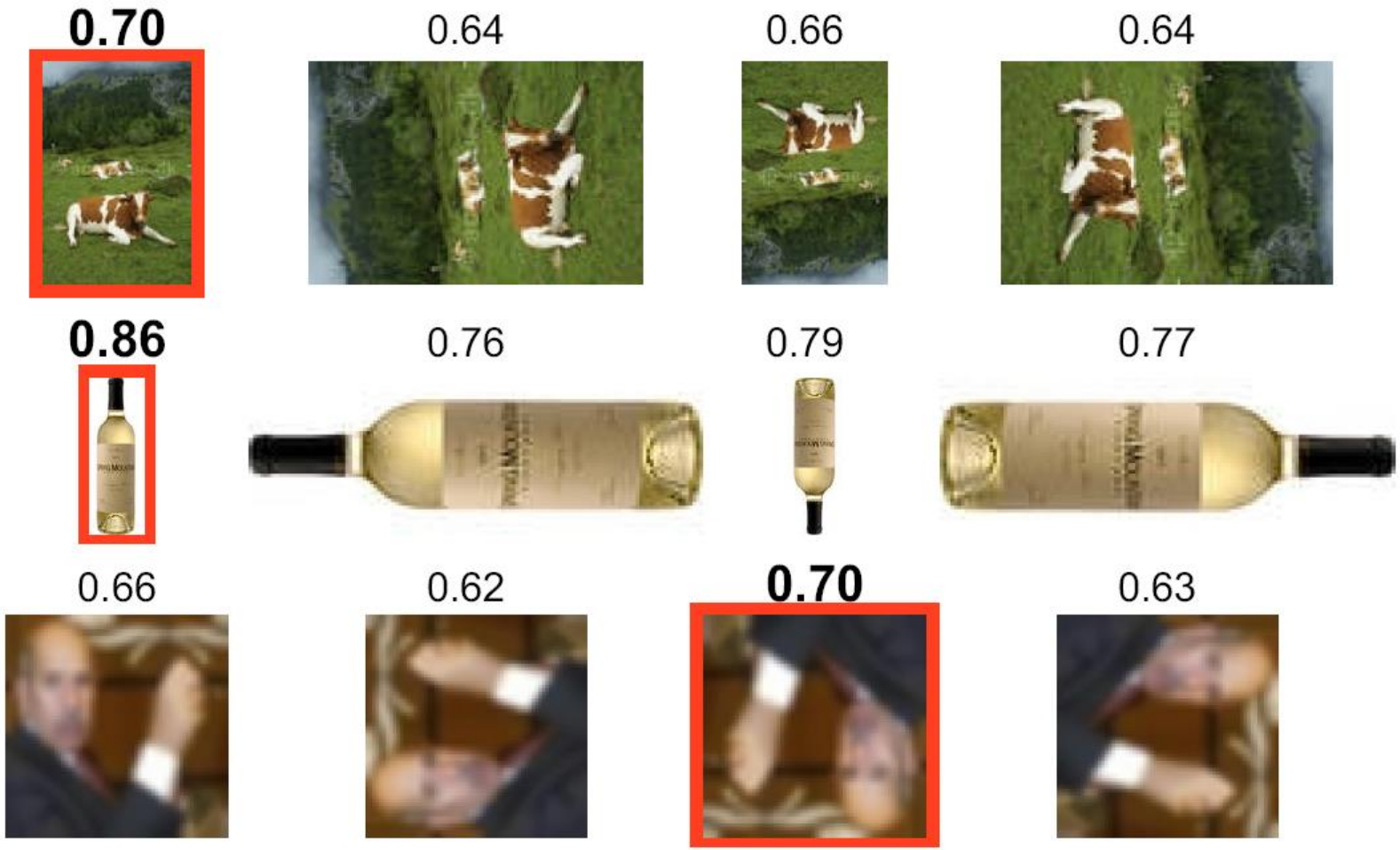
Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match

Automatic Orientation Examples



A. Torralba, R. Fergus, W.T.Freeman. 2008

Summary

- With billions of images on the web, it's often possible to find a close nearest neighbor

- In such cases, we can shortcut hard problems by "looking up" the answer, stealing the labels from our nearest neighbor. For example, simple (or learned) associations can be used to synthesize background regions, colorize, or recognize objects

- But we can't really "brute force" computer vision. Still, it's nice to get an intuition for the size of "image space".

16

Data Sets and Crowdsourcing

Or: My grad students are starting to hate me, but it looks like we need more training data.

Computer Vision

James Hays

The Internet has some rough edges

- https://en.wikipedia.org/wiki/Tay_(bot) in 2016

Microsoft was "deeply sorry for the unintended offensive and hurtful tweets from Tay", and would "look to bring Tay back only when we are confident we can better anticipate malicious intent that conflicts with our principles and values".

June 29th, 2020

It has been brought to our attention [1] that the Tiny Images dataset contains some derogatory terms as categories and offensive images. This was a consequence of the automated data collection procedure that relied on nouns from WordNet. We are greatly concerned by this and apologize to those who may have been affected.

The dataset is too large (80 million images) and the images are so small (32 x 32 pixels) that it can be difficult for people to visually recognize its content. Therefore, manual inspection, even if feasible, will not guarantee that offensive images can be completely removed.

We therefore have decided to formally withdraw the dataset. It has been taken offline and it will not be put back online. We ask the community to refrain from using it in future and also delete any existing copies of the dataset that may have been downloaded.

How it was constructed: The dataset was created in 2006 and contains 53,464 different nouns, directly copied from Wordnet. Those terms were then used to automatically download images of the corresponding noun from Internet search engines at the time (using the available filters at the time) to collect the 80 million images (at tiny 32x32 resolution; the original high-res versions were never stored).

Why it is important to withdraw the dataset: biases, offensive and prejudicial images, and derogatory terminology alienates an important part of our community -- precisely those that we are making efforts to include. It also contributes to harmful biases in AI systems trained on such data. Additionally, the presence of such prejudicial images hurts efforts to foster a culture of inclusivity in the computer vision community. This is extremely unfortunate and runs counter to the values that we strive to uphold.

Yours Sincerely,

Antonio Torralba, Rob Fergus, Bill Freeman.

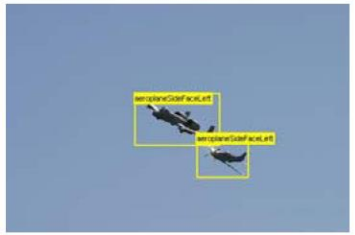
[1] Large image datasets: A pyrrhic win for computer vision?, anonymous authors, OpenReview Preprint, 2020.

Outline

- Data collection with experts – PASCAL VOC
- Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive

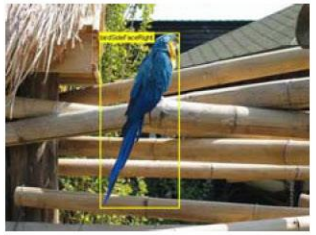
Examples

Aeroplane

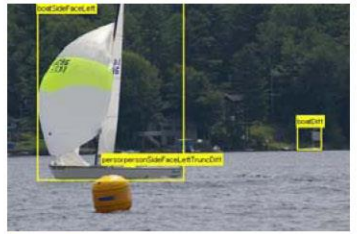


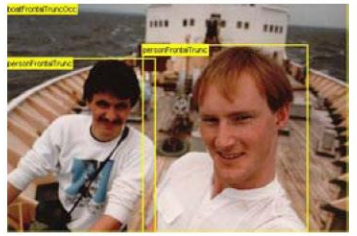
Bicycle

Bird



Boat





Bottle

Bus

Car

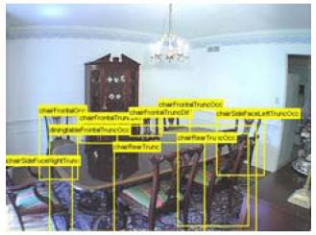
Cat

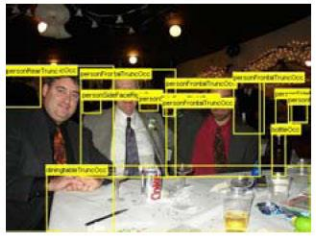
Chair

Cow

Examples

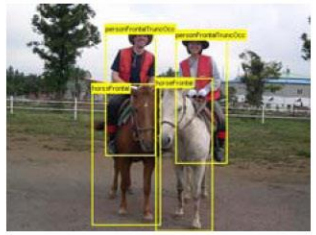
Dining Table



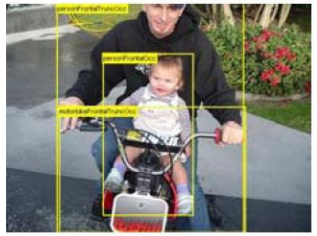


Dog

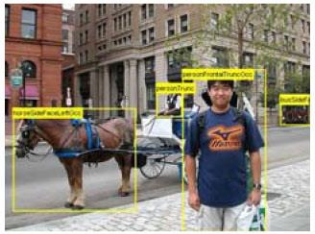
Horse

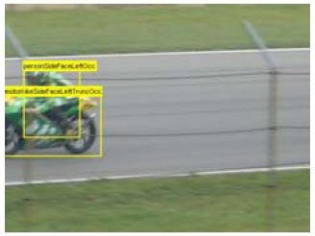


Motorbike

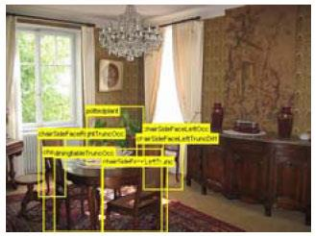


Person

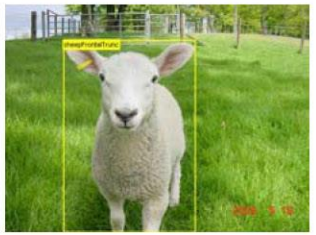




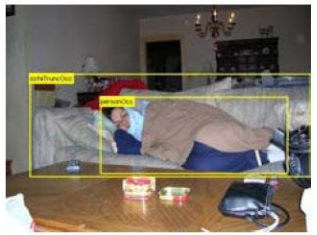
Potted Plant



Sheep

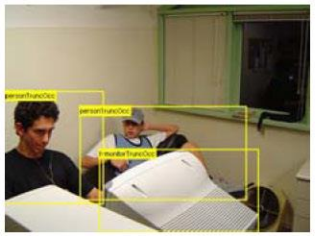


Sofa



Train

TV/Monitor



Outline

- Data collection with experts – PASCAL VOC
- Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive

LabelMe

- http://labelme.csail.mit.edu
- “Open world” database annotated by the community*

- * **Notes on Image Annotation**, Barriuso and Torralba 2012. http://arxiv.org/abs/1210.3448

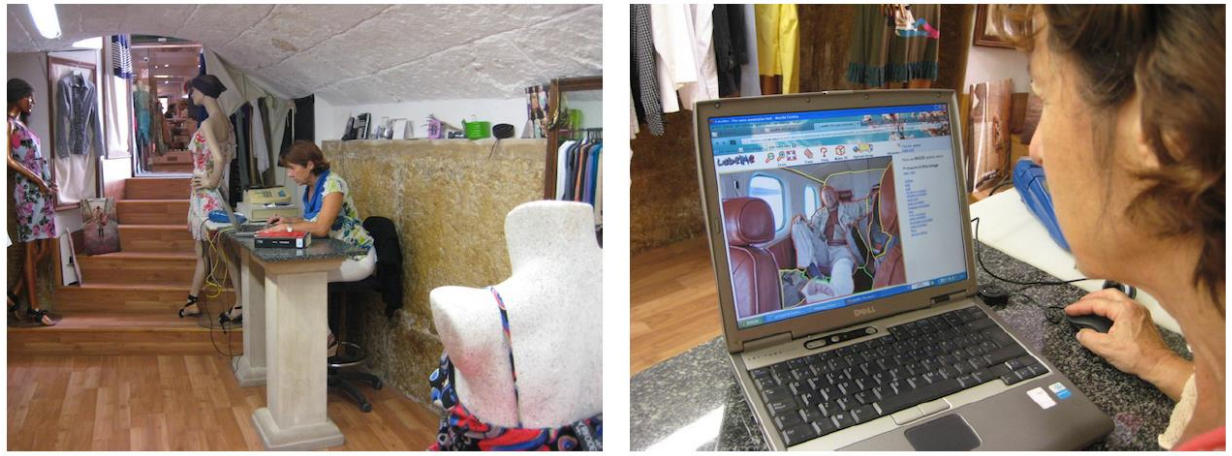


Figure 2: *The image annotation context. All the labeling was done inside a clothing shop named* Transparencia *in the heart of Palma de Mallorca, Spain.*

knowledge of typical contextual arrangements?

It is often said that vision is effortless, but frequently the visual system is lazy and makes us believe that we understand something when in fact we don't. In occasions we find ourselves among objects whose names and even functions we may not know but we do not seem to be bothered by this semantic blindness. However, this changes when we are labeling images as we are forced to segment and name all the objects. Suddenly, we are forced to see where our semantic blind-spot is. We become aware of gaps in our visual understanding of what is around us.

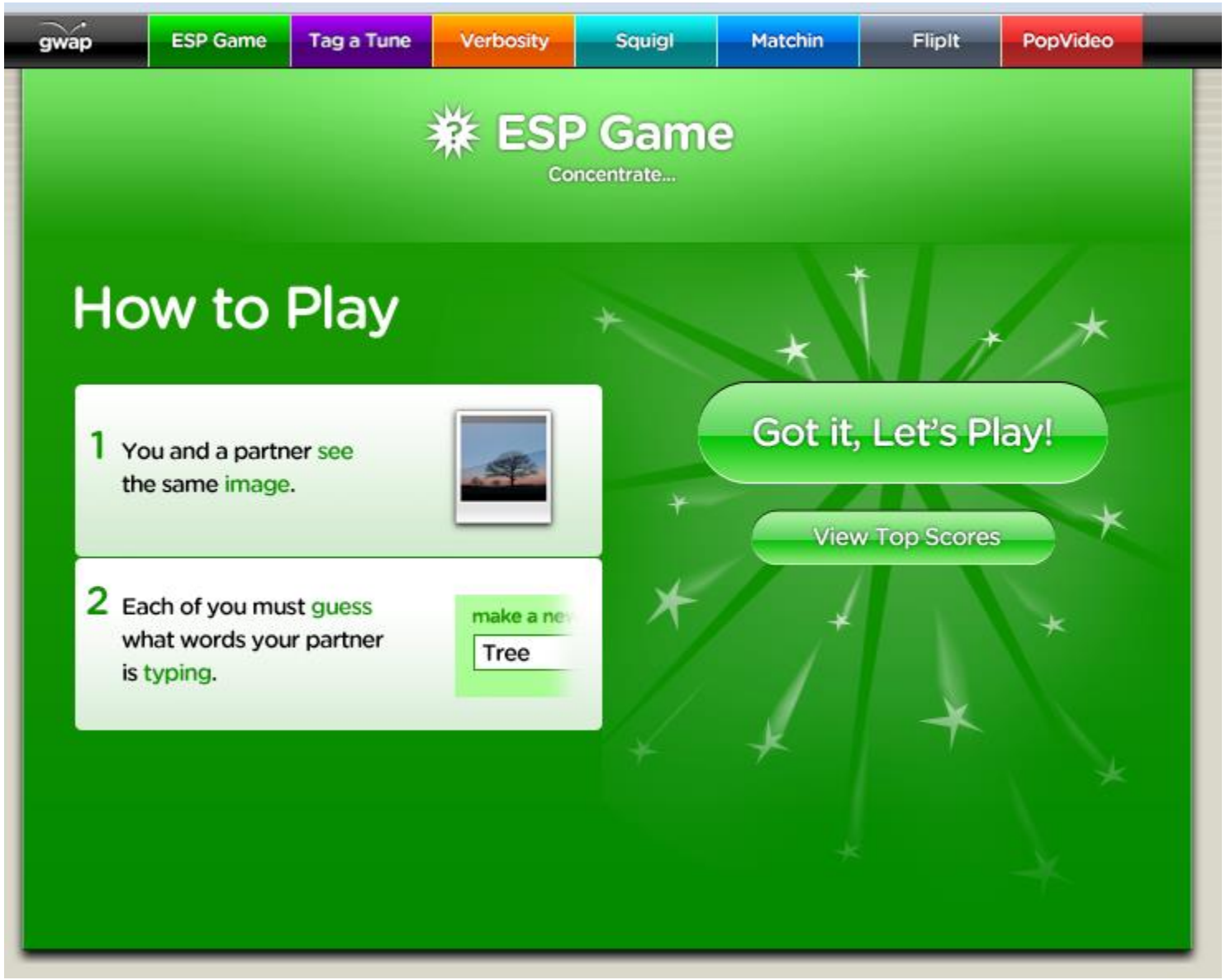
This paper contains the notes written by Adela Barriuso describing her experience while using the LabelMe annotation tool [1]. Since 2006 she has been frequently using LabelMe. She has no training in computer vision. In 2007 she started to use LabelMe to systematically annotate the SUN database [7]. The goal was to build a large database

there is not a fix set of categories. As the goal is to label all the objects within each image, the list of categories grows unbounded. Many object classes appear only a few times across the entire collection of images. However, not even those rare object categories can be ignored as they might be an important element for the interpretation of the scene. Labeling in these conditions becomes difficult as it is important to keep a list of all the object classes in order to use a consistent set of terms across the entire database avoiding synonyms. Despite the annotator best efforts, the process is not free of noise.

Since she started working with LabelMe, she has labeled more than 250,000 objects. Labeling more than 250,000 objects gives you a different perspective on the act of seeing. After a full day of labeling images, when you walk on the street or drive back home, you see the world in a different way. You see polygons outlining objects, you

"Since she started working with LabelMe, she has labeled more than 250,000 objects."

Notes on Image Annotation, Barriuso and Torralba 2012. http://arxiv.org/abs/1210.3448



Luis von Ahn and Laura Dabbish. Labeling Images with a Computer Game.
ACM Conf. on Human Factors in Computing Systems, CHI 2004

Outline

- Data collection with experts – PASCAL VOC
- Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive

ESP Game | Tag a Tune | Verbosity | Squigl | Matchin | logged i

score
0

ESP Game
Concentrate...

time
2:56

What do you see?

taboo words
student

guesses

submit | pass

Play Anonymously

Screencast-O-Matic.com

Outline

- Data collection with experts – PASCAL VOC
- Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive

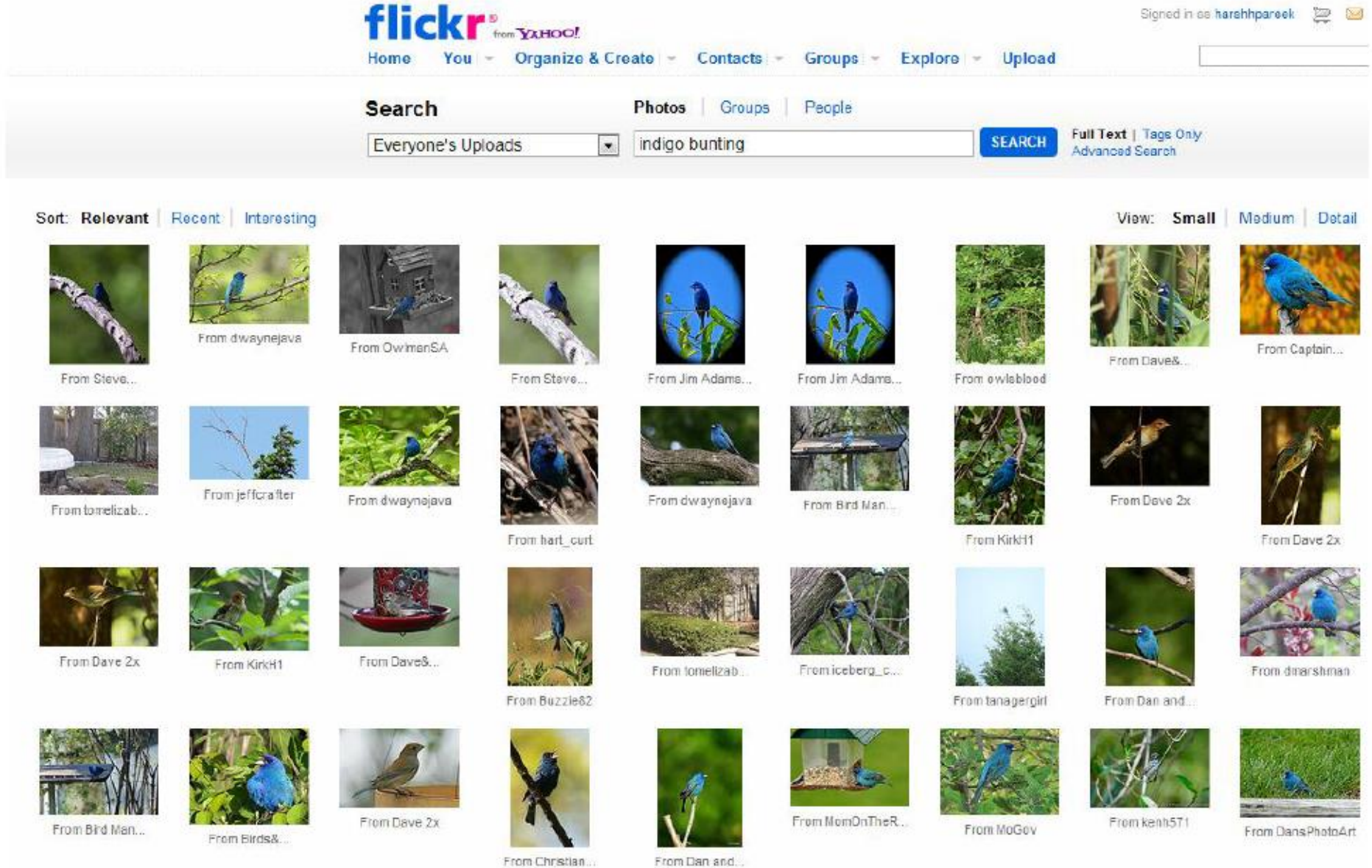
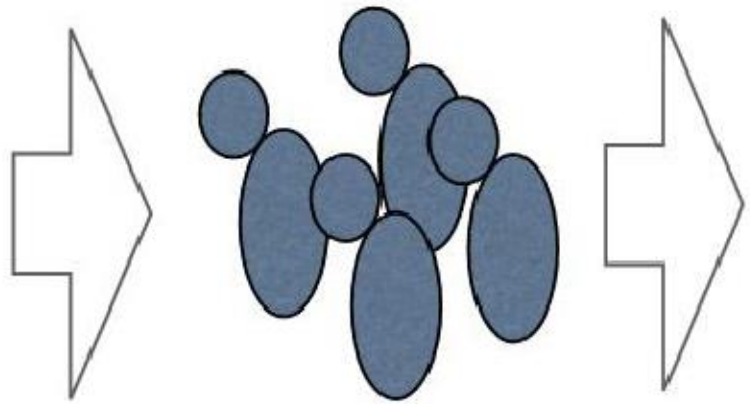


Image credit: Flickr.com

Building datasets

6000 images from flickr.com

Annotators



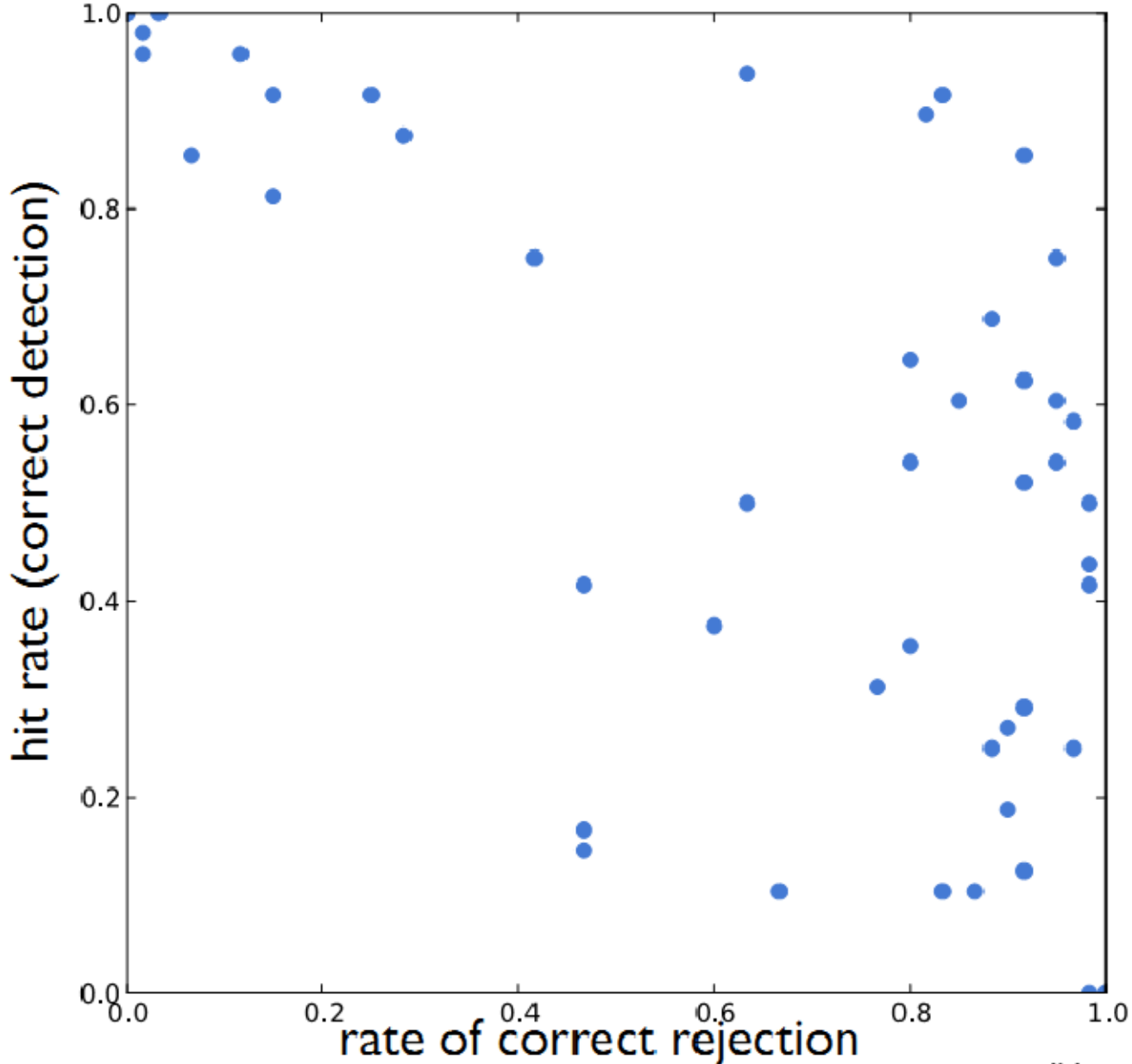
amazonmechanical turk
Artificial Artificial Intelligence

Is there an Indigo bunting in the image?

100s of training images

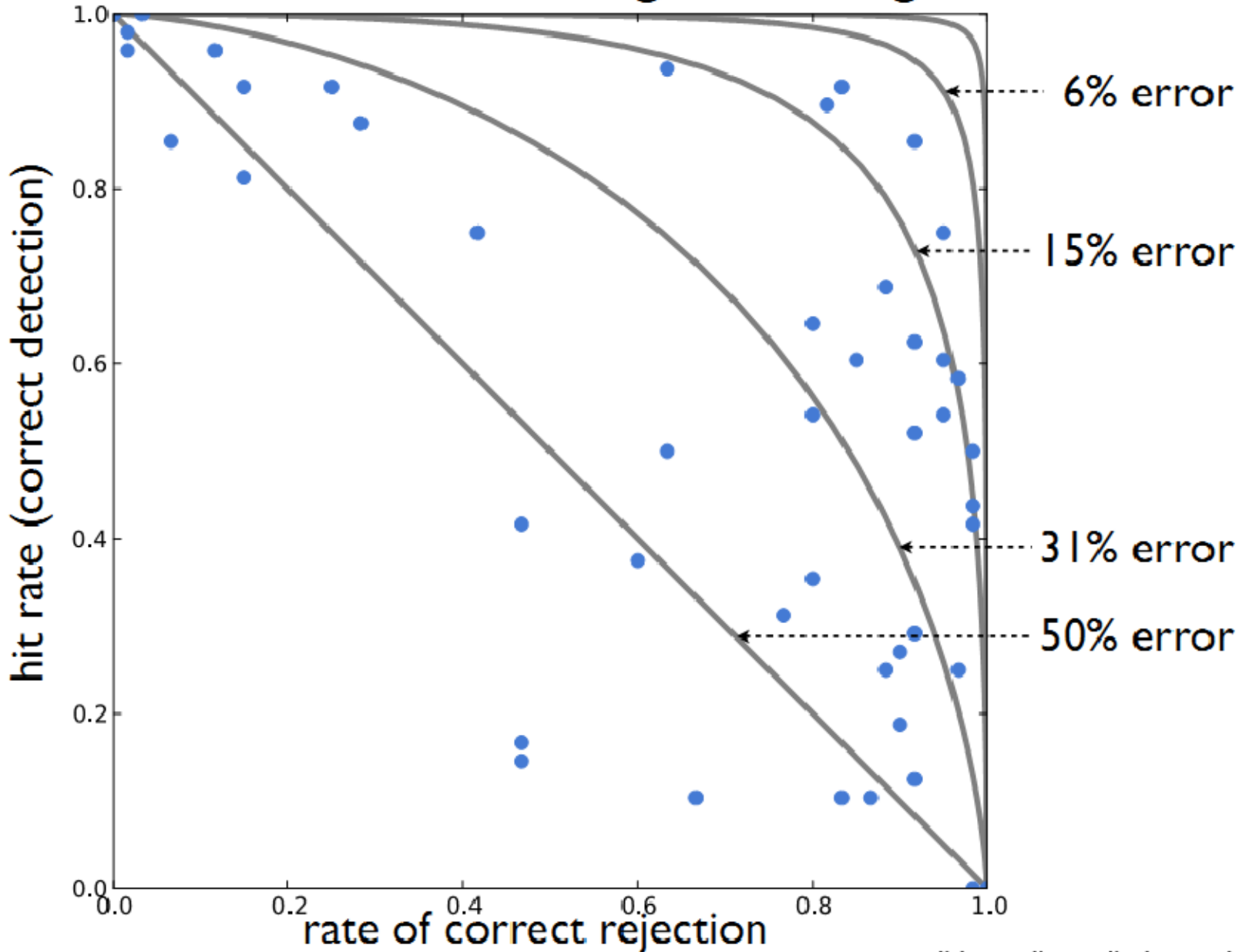
Slide credit: Welinder et al

Task: Find the Indigo Bunting



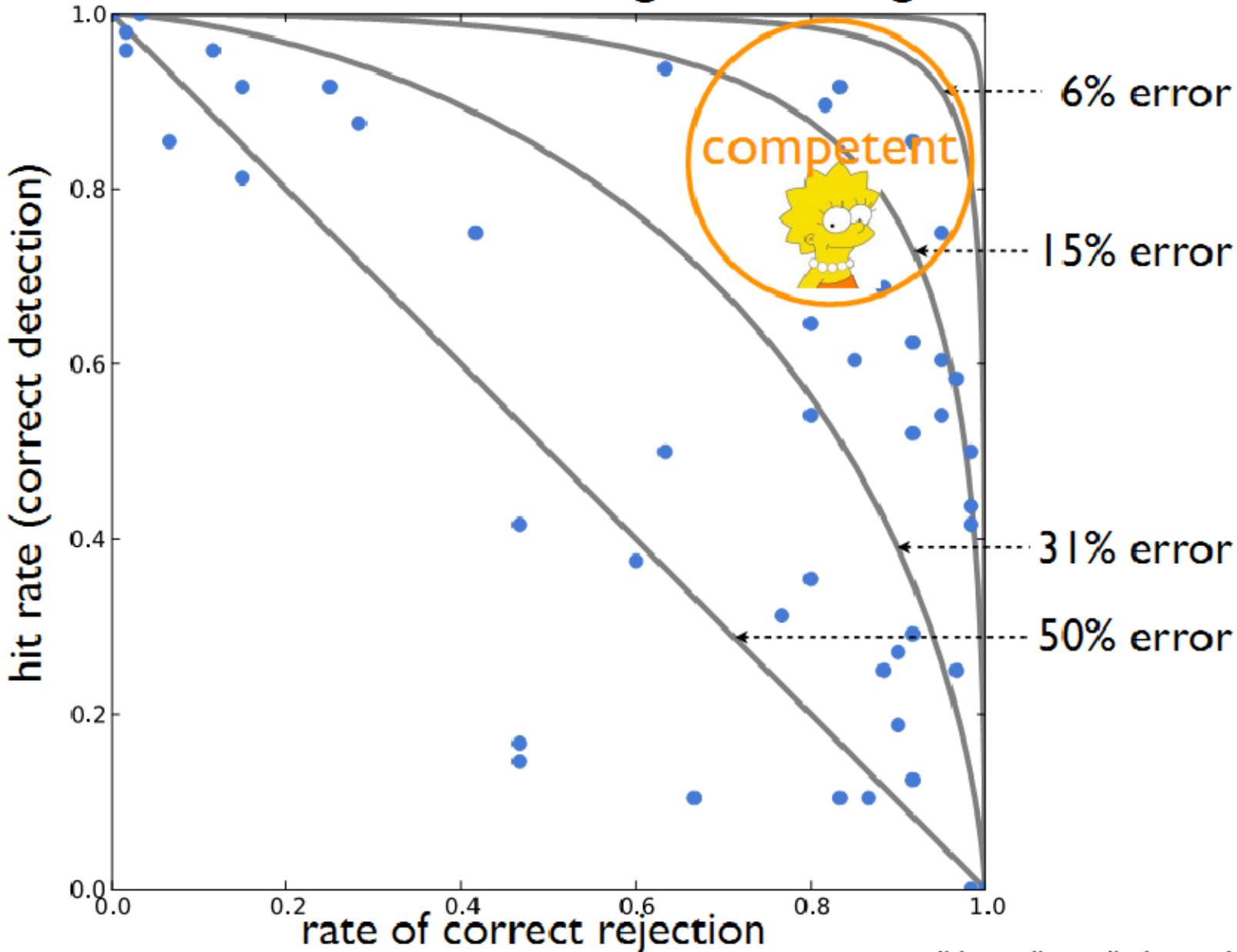
Slide credit: Welinder et al

Task: Find the Indigo Bunting



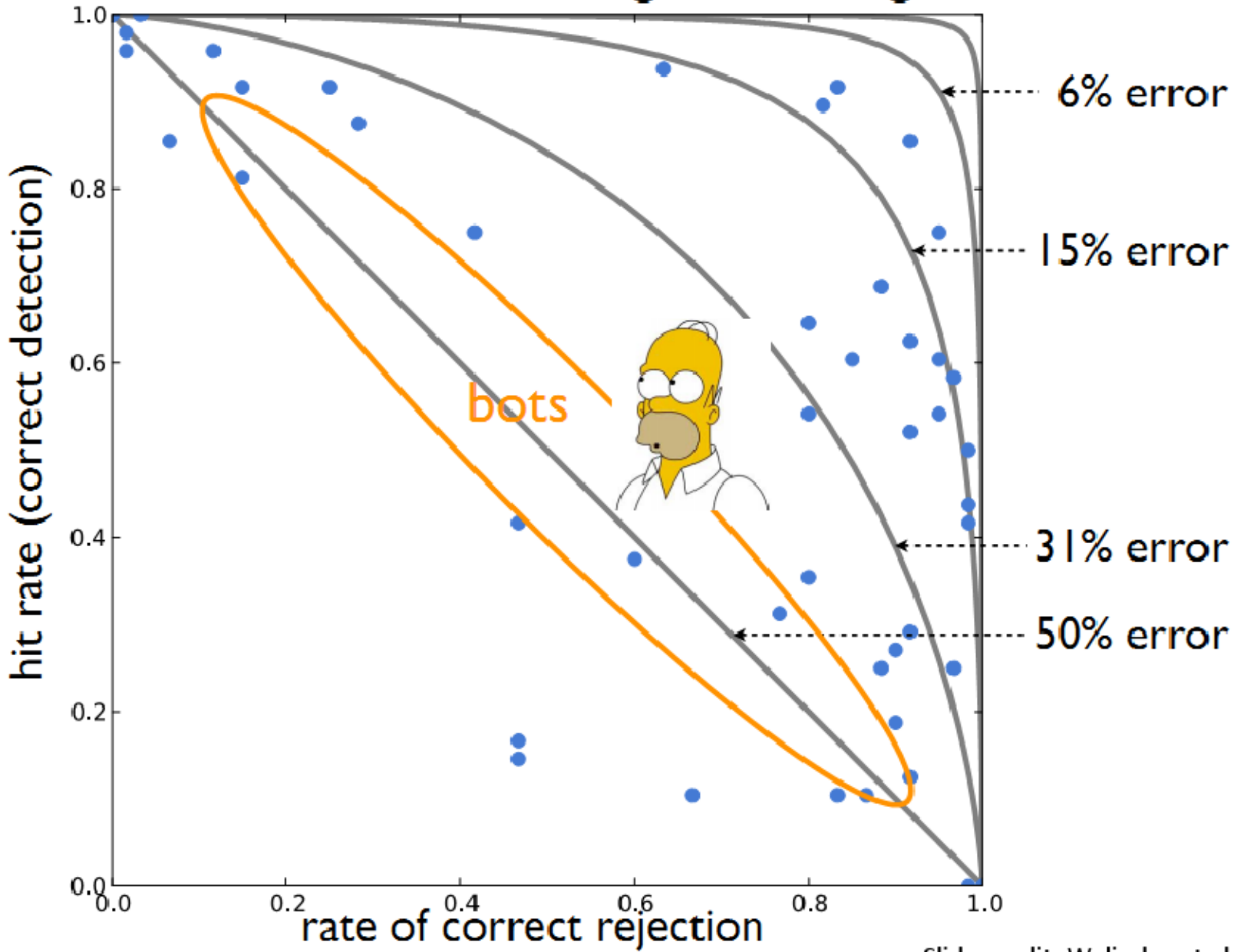
Slide credit: Welinder et al

Task: Find the Indigo Bunting



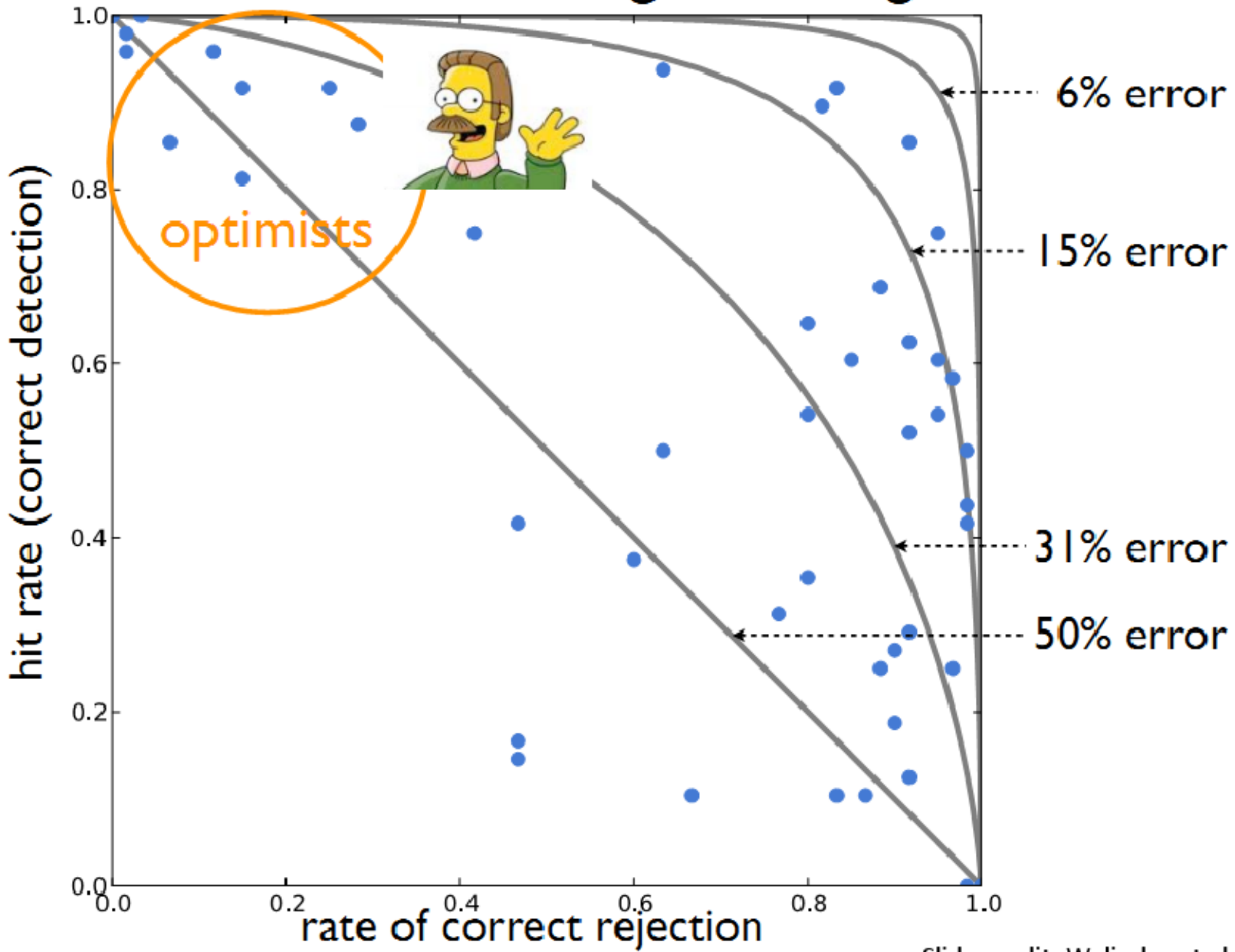
Slide credit: Welinder et al

Task: Find the Indigo Bunting



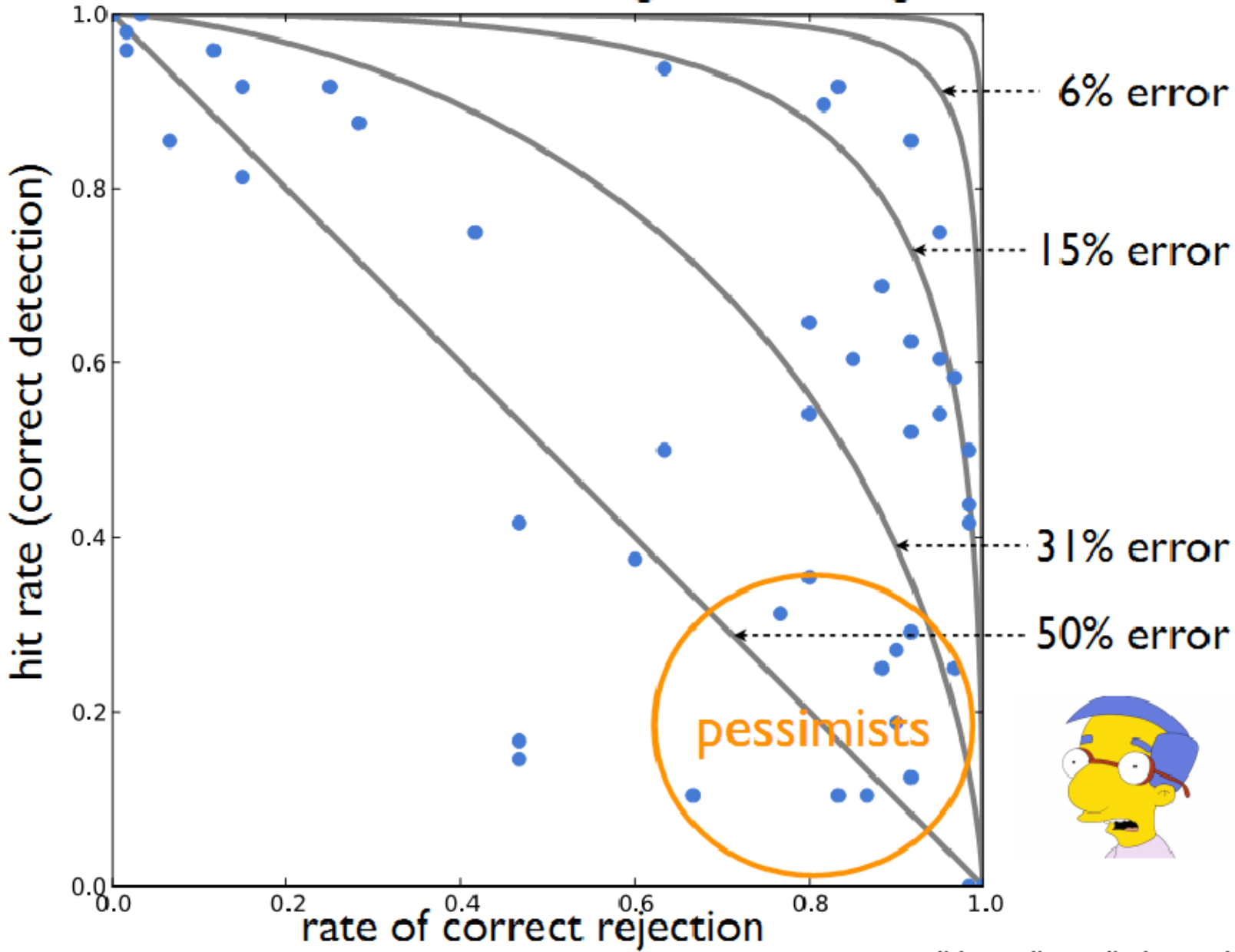
Slide credit: Welinder et al

Task: Find the Indigo Bunting



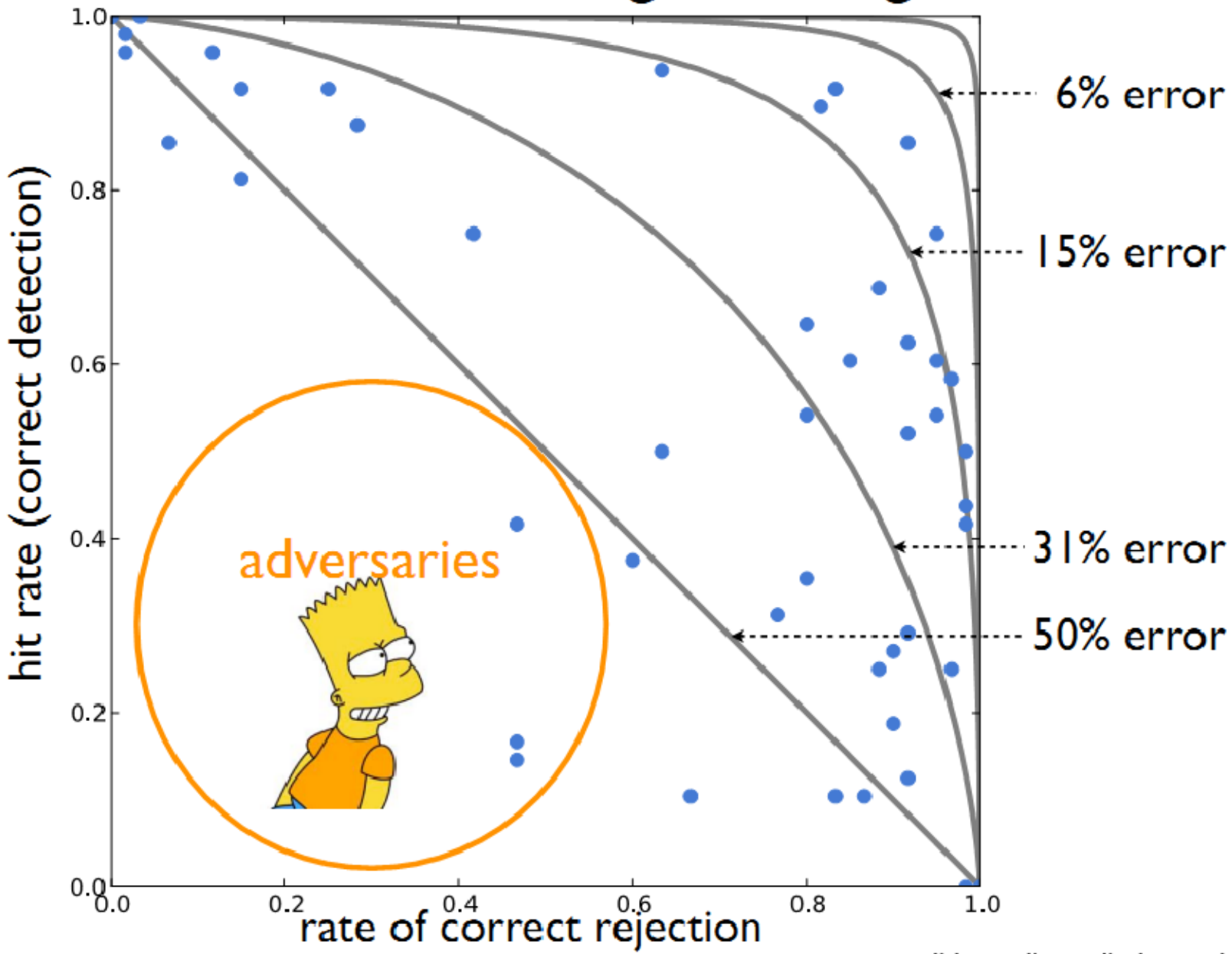
Slide credit: Welinder et al

Task: Find the Indigo Bunting



Slide credit: Welinder et al

Task: Find the Indigo Bunting



Slide credit: Welinder et al

Utility data annotation via Amazon Mechanical Turk

X 100 000 = $5000

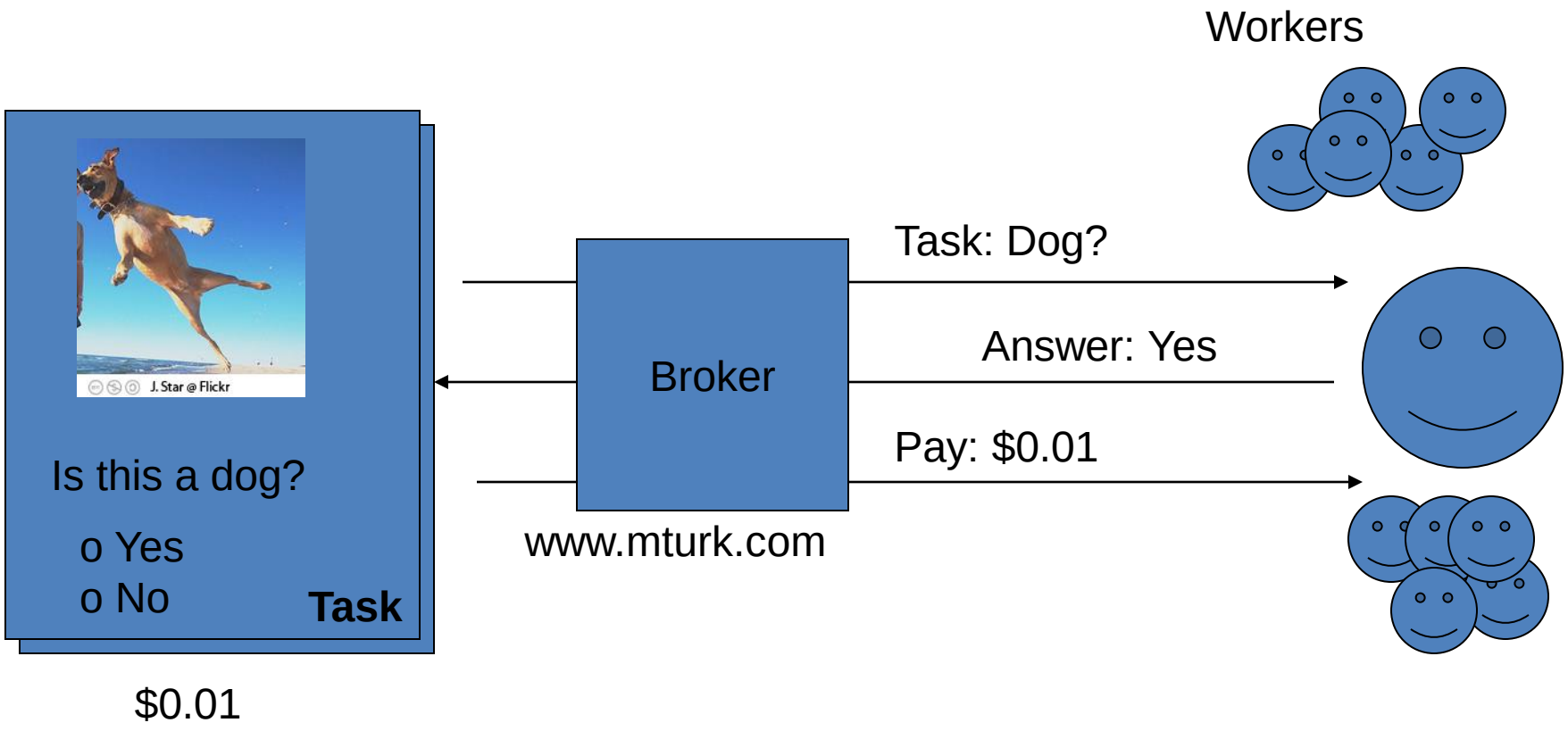
Alexander Sorokin

David Forsyth

CVPR Workshops 2008

Slides by Alexander Sorokin

Amazon Mechanical Turk



Annotation protocols

- Type keywords
- Select relevant images
- Click on landmarks
- Outline something
- Detect features

…….. anything else ……

Type keywords

Mechanical Turk Project

If you're using the turk, Be sure to copy the text back into the HIT page so that you can be credited.

- Photo should be rotated 90 degrees left (counter-clockwise)
- Photo should be rotated 90 degrees right (clockwise)
- Photo should be turned upside down
- Photo is oriented properly

Please describe the picture in the box using 10 words or more:

shells

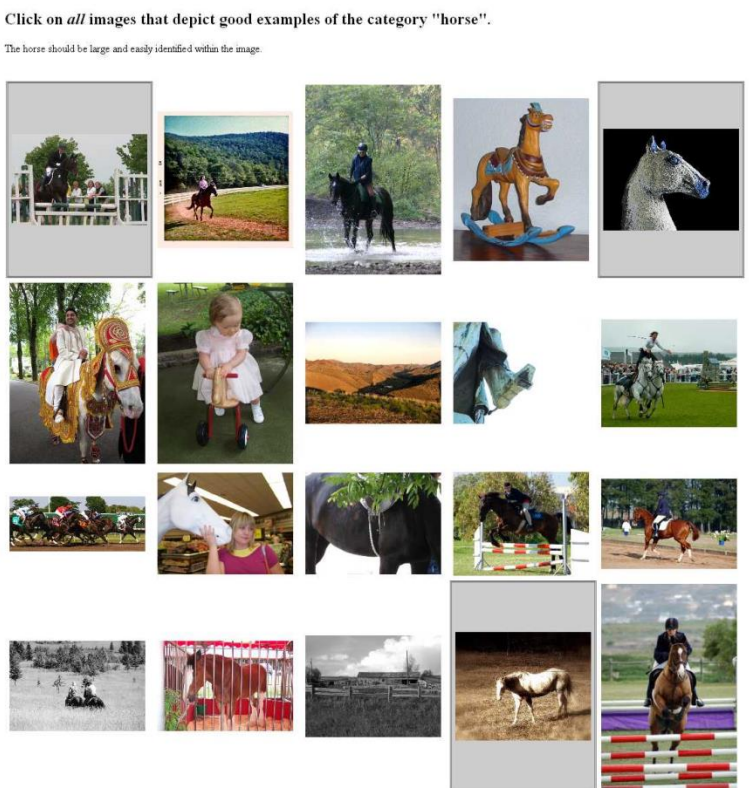
Submit Turk | Skip / Load a different photo

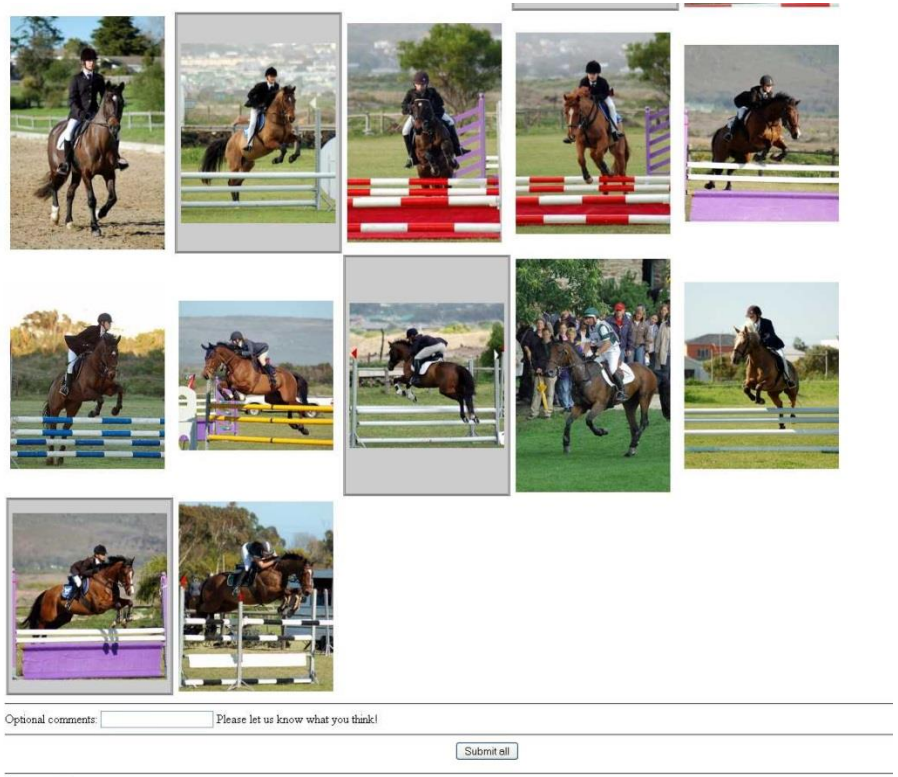
The submit button MUST be clicked!

$0.01

http://austinsmoke.com/turk/.

Select examples

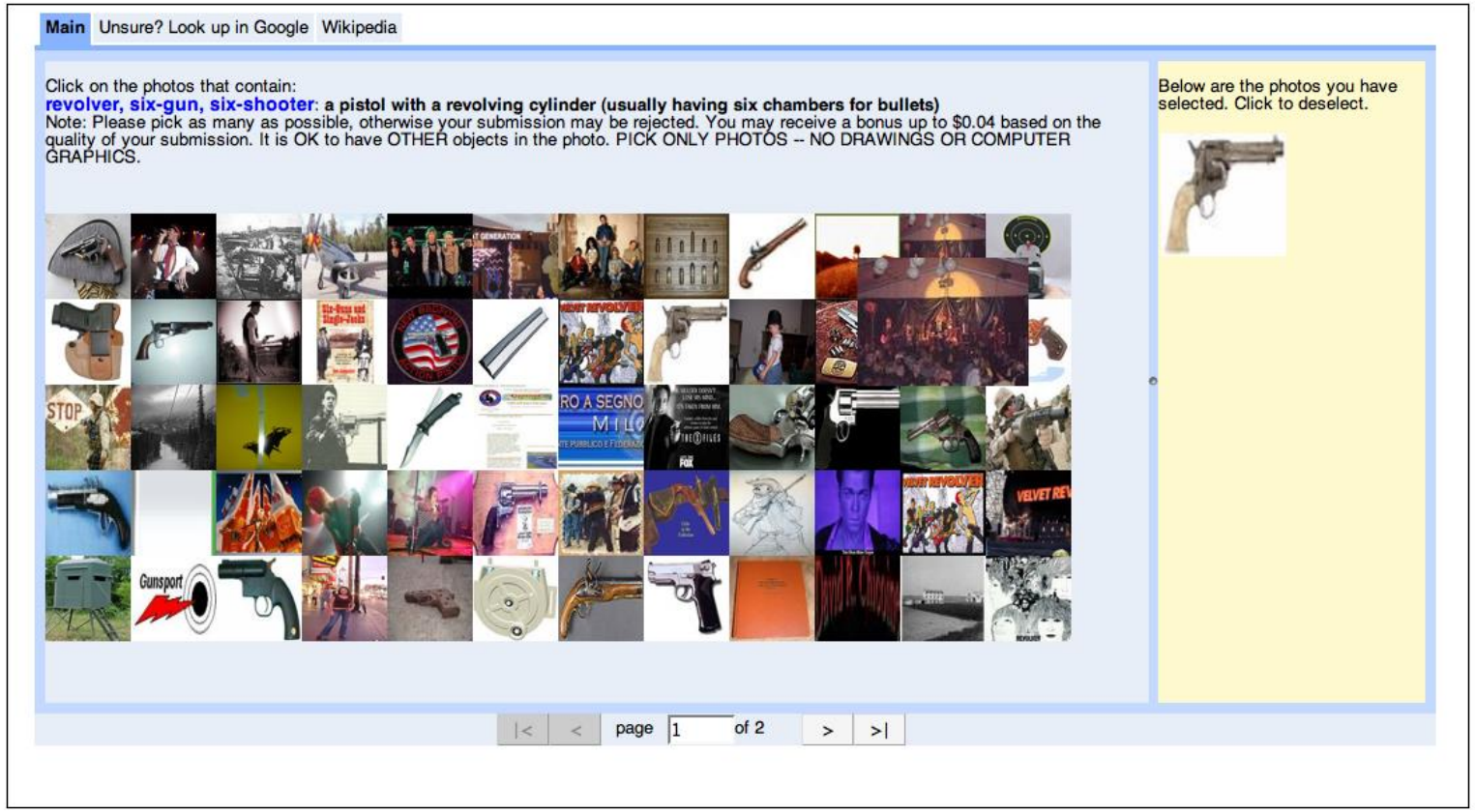




Joint work with Tamara and Alex Berg

http://visionpc.cs.uiuc.edu/~largescale/data/simpleevaluation/html/horse.html

Select examples



$0.02

requester mtlabel

Click on landmarks

$0.01

http://vision-app1.cs.uiuc.edu/mt/results/people14-batch11/p7/

Outline something

$0.01

http://visionpc.cs.uiuc.edu/~largescale/results/production-3-2/results_page_013.html

Data from Ramanan NIPS06

Motivation

Custom annotations

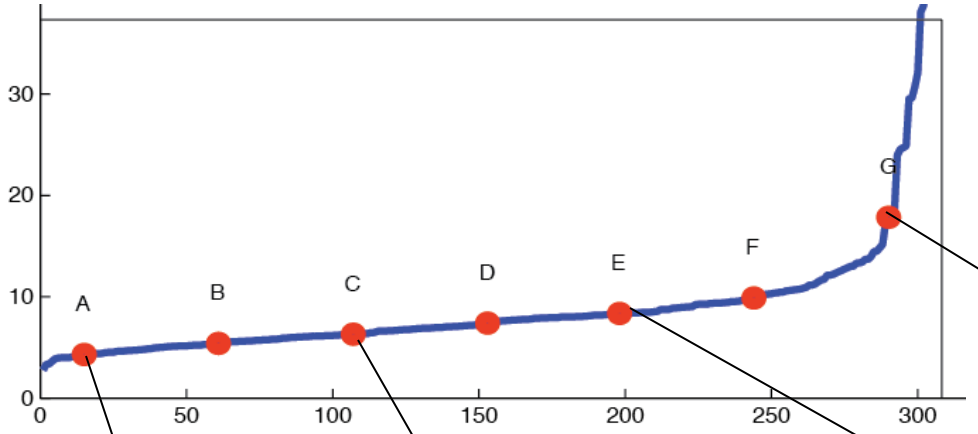
X 100 000 Large scale

= $5000 Low price

Issues

- Quality?
  - How good is it?
  - How to be sure?
- Price?
  - How to price it?

Annotation quality



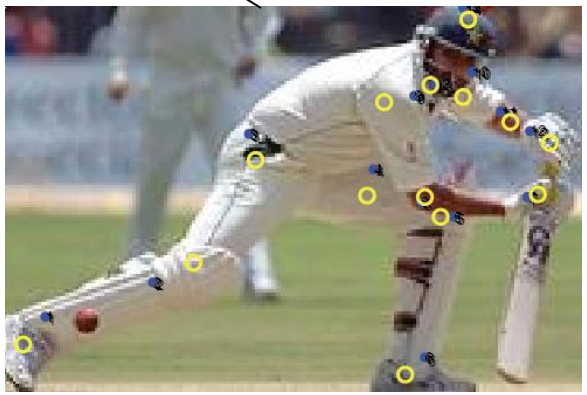
Agree within 5-10 pixels on 500x500 screen

There are bad ones.

A

C

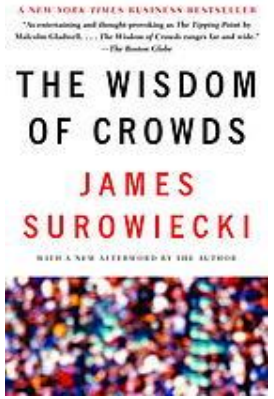
E



G

How do we get quality annotations?

Ensuring Annotation Quality



- Consensus / Multiple Annotation / "Wisdom of the Crowds"

  Not enough on its own, but widely used

- Gold Standard / Sentinel
  - Special case: qualification exam

    Widely used and most important. Find good annotators and keep them honest.
- Grading Tasks
  - A second tier of workers who grade others

    Not widely used

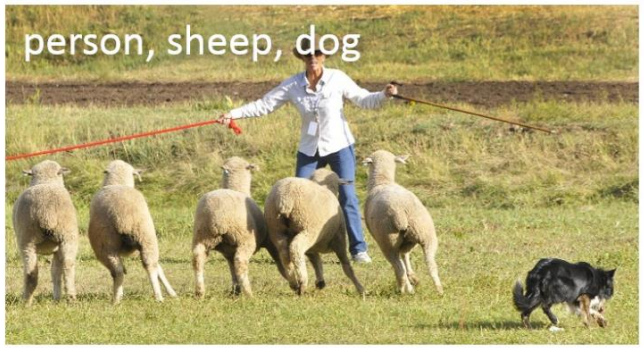
Pricing

- Trade off between throughput and cost
  - *NOT* as much of a trade off with quality
- Higher pay can actually attract scammers

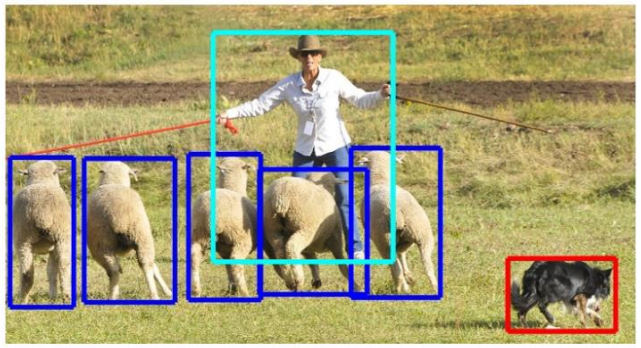
Examples of Crowdsourcing

- Massive annotation efforts that would not otherwise be feasible
  - ImageNet (http://www.image-net.org/)
  - COCO (http://cocodataset.org)
  - Many more

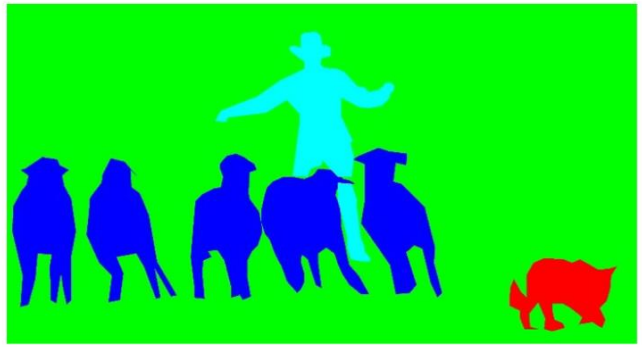
Crowdsourcing to build MS COCO



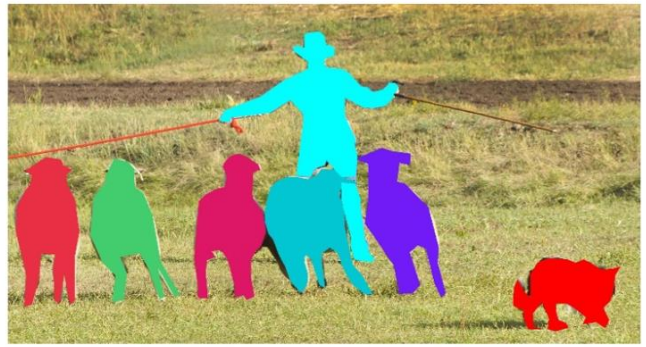
(a) Image classification



(b) Object localization



(c) Semantic segmentation



(d) This work

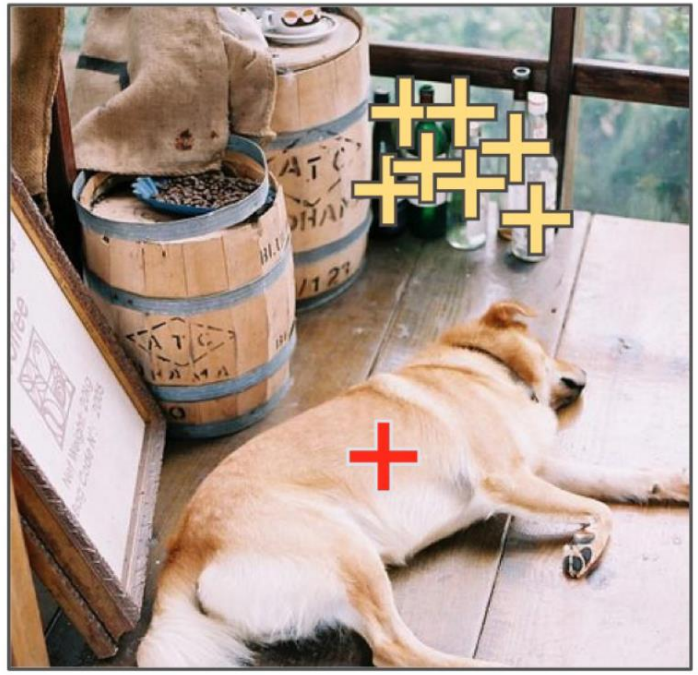
Microsoft COCO: Common Objects in Context

Tsung-Yi Lin Michael Maire Serge Belongie Lubomir Bourdev Ross Girshick
James Hays Pietro Perona Deva Ramanan C. Lawrence Zitnick Piotr Dollár

Crowdsourcing to build MS COCO

Annotation Pipeline

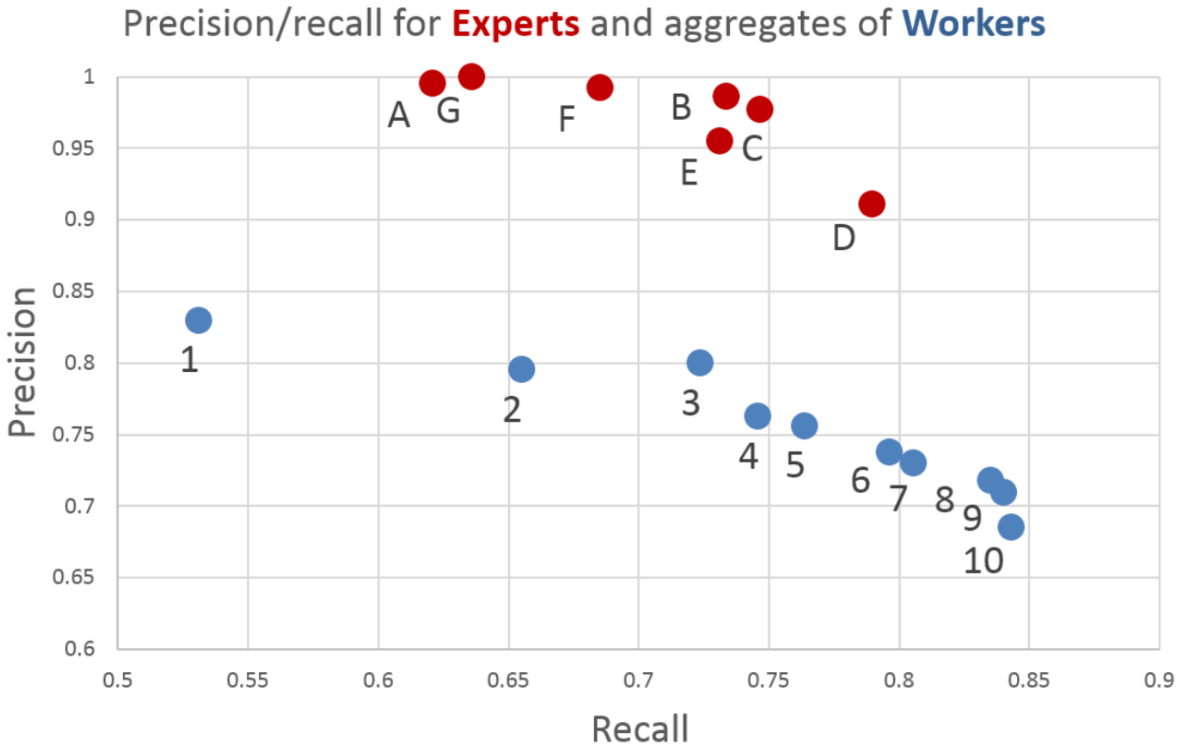
(a) Category labeling

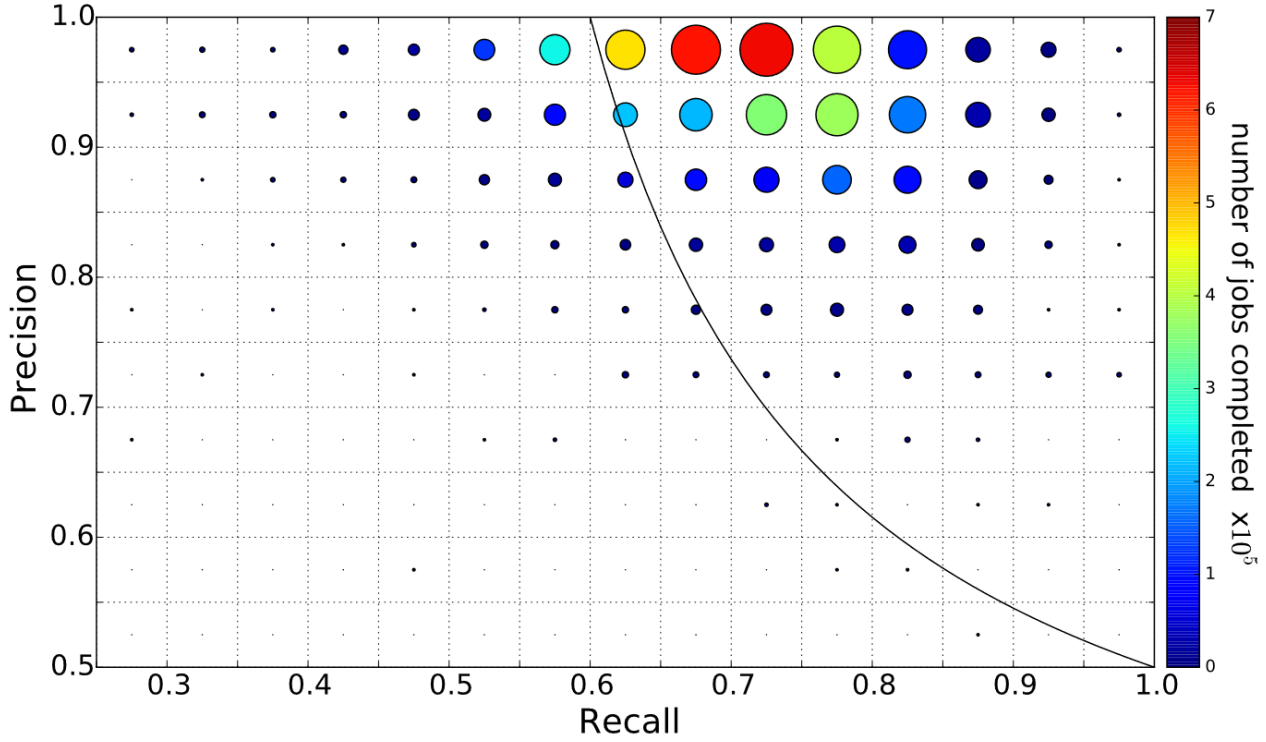


(b) Instance spotting

(c) Instance segmentation

Crowdsourcing to build MS COCO





Examples of Crowdsourcing

- Most papers annotate images, but there are some more creative uses
  - Webcam Eye tracking (https://webgazer.cs.brown.edu/)
  - Sketch collection (http://cybertron.cg.tu-berlin.de/eitz/projects/classifysketch/)
    - Flips the usual annotation process, by providing a *label* and asking for an *image*

Next lecture

- "Unsupervised" Deep Learning